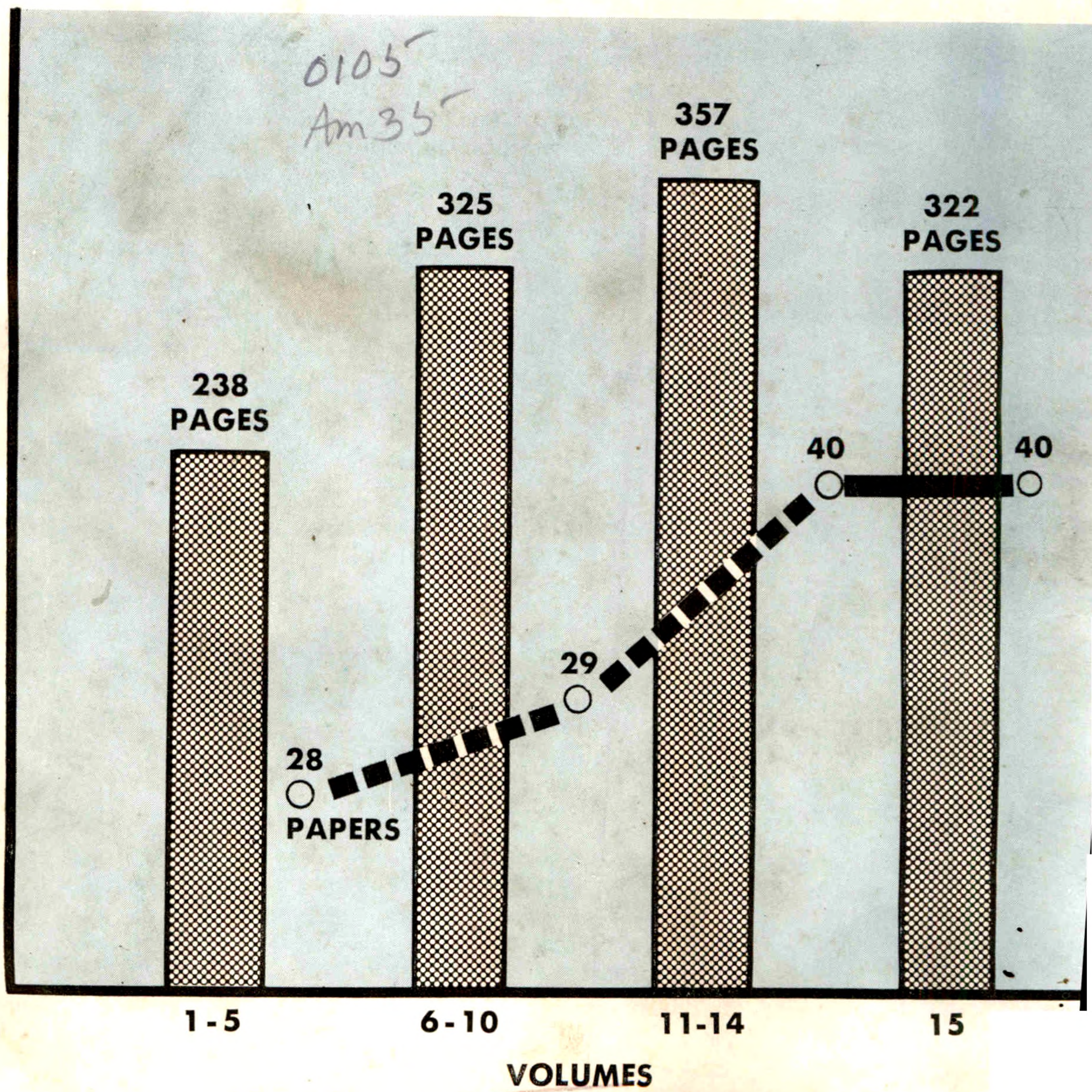
American Documentation

JANUARY

1965

Vol. 16, No. 1

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE



AMERICAN DOCUMENTATION

American Documentation is a publication of the American Documentation Institute. Its purpose is to be a scholarly journal in the various fields of documentation and to serve as a forum of discussion and experimentation.

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), one copy in the institutional membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

Instructions to authors are given in Volume 14, No. 2 (April 1963). All manuscripts, brief communications (1000 words or less), announcements, or letters should be sent to:

Mr. Arthur W. Elias, Editor
American Documentation
Institute for Scientific Information
325 Chestnut Street
Philadelphia, Pennsylvania 19106

Correspondence on reviews should be addressed to:
Dr. Theodore C. Hines, Review Editor
American Documentation
54 North Drive
East Brunswick, New Jersey

Correspondence on Literature Notes should be addressed to:
Mr. Charles P. Bourne, Literature Notes Editor
American Documentation
Stanford Research Institute
Menlo Park, California

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts,* and *Science Citation Index.*

American Documentation is entered for second class mailing at Washington, D. C.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 16, No. 1 JANUARY 1965

Published quarterly by American Documentation Institute and printed by Spartan Books, Inc., 1106 Connecticut Avenue, Washington, D. C. Second class postage paid at Washington, D. C.

Obituary

G. MILES CONRAD (1911-1964)

The Editor regretfully must report the passing of G. Miles Conrad, Director of Biological Abstracts, and valued member and friend of the American Documentation Institute. Mr. Conrad served as Secretary of the Institute from 1951 until 1954, and was internationally known as an expert in the areas of abstracting and biological documentation. He served as President of the National Federation of Science Abstracting and Indexing Services in 1958. ADI extends its sympathies to his family and associates at Biological Abstracts.

Editorial

The Editor's Annual Report to the Institute is reprinted in full to provide a complete accounting to members and subscribers. In addition, the Editor is most pleased to announce the termination of the page charge policy for *American Documentation* for 1965. It is hoped that it will not be necessary to reinstitute this practice in the future. Further, it is now possible to receive Literature Notes as a separate signature. The price for this is set at $4.00 per year.

We would be pleased to receive comments and suggestions.

Editor's Annual Report

AMERICAN DOCUMENTATION has published 40 refereed items* in the four issues for 1964. Of these, 31 were journal articles and 9, Brief Communications—a new publication format initiated in the April issue as a solution to two problems; the time lag for full publication, and the need for an alternate publication form to cover material which for one reason or another is unsuited to journal article format. Non-refereed items included 8 Letters to the Editor, 6 Book Reviews, 70 pages of Literature Notes, 9 pages of a Quadrennial Index, and 8 pages of Contents for Binding purposes.**

Organization and editorial changes for 1964 included the appointment of an Associate Editor to serve as back-up for the Editor in attendance at Council meetings, provide editorial continuity in the event of the Editor's inability to perform his function, and to assure a second level of consultation on editorial matters. In addition, the function of Reviews Editor was established to cover this area in procuring materials for review, etc.

A policy of multiple refereeing was initiated and a reviewing form designed to provide levels of selectivity and accuracy, inform authors of specific problems, and insure a fair editing procedure, has been instituted.

The use of a half-tone on covers, initiated by Mr. J. Mack, has been continued and coordinated with the theme of the editorial in most cases.

Revisions of publishing procedures have been made, facilitating the processing of all issues.

The production of the quadrennial index brings AD up-to-date. It is planned to go on an annual basis from here on in. The thanks of the Editor and the Institute are in order to Mr. Boris Anzlowar for doing this index at no charge.

A policy of reports to the Editorial Board has been maintained to keep them informed of problems and plans for the journal. In addition, this highly skilled and cooperative group has been called on with increasing frequency to aid in the refereeing procedures.

The Editorial Board and Staff of AD took on the responsibility for the production of the 1964 Proceedings of ADI, which will be available at the annual meeting. The production of a 500-page, hard-cover book in a period of 13 weeks was an enormous undertaking, and while there were problems and possibly errors of omission and commission, for the most part tremendous cooperation and efficiency was forthcoming.

Thirty-six refereed items are on hand for 1965. Of these, 30 are journal articles and 6 are Brief Communications. Since the target for the latter is publication within 6 months, some details are necessary. Three of the six remain from 1963 and were originally submitted as journal articles. All have been refereed and returned to the authors for revision. Publication has not occurred due to the fact that authors have retained the manuscripts and have not resubmitted or withdrawn the items. The remaining three items are still within the 6 month time period and will be forthcoming in 1965.

Of the 30 journal articles, 9 date from 1963, with 21 remaining from 1964. The 1963 items are in much the same status as the 1963 Brief Communications.

Average time from receipt to publication for 1964 was 10 months. 20% of the items were published in 6 months or less, 25% in 12 months or more. The maximum lags were 16 and 15 months respectively, each occurring once.

ARTHUR W. ELIAS, *Editor*
AMERICAN DOCUMENTATION

* *Refereed items* defined as material submitted for technical review.

** Contents pages for 1963 were omitted in the October issue for that year and had to be provided in the 1964 issue. Normally, pages are required in any given year.

Some Clumping Experiments for Associative Document Retrieval

The paper describes a computational procedure for defining associations between words used to index a collection of documents. Using association structures generated by the procedure in a small document collection, a retrieval model is described and the results of trial retrievals are evaluated. The techniques described appear to be applicable in practice for retrieval in large collections.

A. G. DALE and N. DALE

Linguistics Research Center
The University of Texas
Austin, Texas

This paper describes experiments with a statistical technique that can be used to compute word associations useful for document retrieval purposes in a collection where documents are described by index words occurring in the text or in abstracts of the documents.

Although the experiments have been conducted with a small collection of documents, they have suggested methods that will permit the association technique to be applied in collections of an interesting size. This is important, since a number of alternative methods that have been proposed for computing word associations appear to present intractable computation problems that will deter large scale application.

• Why Associative Retrieval?

There are three principal reasons why associative retrieval techniques are worth investigating:

1. The efficiency of retrieval in a coordinate indexed system rapidly diminishes as the document collection grows in size. In an associative retrieval system, on the other hand, retrieval efficiency as measured by recall and relevance ratios will not necessarily decrease with the size of the document collection, and might, in fact, improve (1).
2. If the structure of word associations in a given collection is adequate, the user of the collection can define a retrieval request in his own terms without reference to an elaborate dictionary or thesaurus of index terms. Thus, the need for interposing an interpreter between the user and the collection can be dispensed with, and the exponentially increasing difficulty of compiling an external thesaurus is avoided.
3. The techniques can be applied to word data that has been automatically processed from document texts, or identified in texts by individuals inexperienced in complex conventional indexing methods, thus permitting rapid indexing and acquisition of materials in a collection.

• Clumping Techniques and Word Associations

The method used to define word associations is based on the theory of clumps (2,3,4,5), a technique developed largely by R. M. Needham of the Cambridge Language Research Unit. The notion of a clump is intuitively simple: a clump is a group of objects from some universe of objects such that the members of the group (technically, the *subset*) are more closely related to each other than to the rest of the universe to which they belong.

In order to find clumps of objects, two decisions must first be made:

1. We have to specify how associations between objects are to be measured.
2. We have to specify the criteria that define a clump.

1. *Measuring associations between index words*

In the context of a set of documents and a set of index words, a classification array can be constructed such that the elements are 1 or 0 depending on whether or not a word is used in a given document.

If we let 1(n) and 1(m) be the number of 1's in rows n and m respectively of an index word-document classifi-

* Work described in this paper was supported in part by the National Science Foundation under Institutional Grant GU-483 at the University of Texas.

cation array, then 1(n) specifies the number of documents in which the n-th word occurs, and 1(m) the number of documents in which the m-th word occurs. Additionally, let 1(n,m) be the number of documents in which the two words co-occur. With these data, a number of similarity measures between pairs of words can be constructed. Two possible definitions of the nearness (or *connection*) between the n-th and m-th terms are:

$$\text{Def. 1:}\quad 1(n,m)$$

$$\text{Def. 2:}\quad \frac{1(n,m)}{\sqrt{1(n)\cdot 1(m)}}$$

The first definition measures nearness between two words on the basis of the simple property of their co-occurrence in documents. The second definition measures the connection between a pair of items by taking into account the conditional probability of co-occurrence, given the total occurrences of each item.

Having selected a particular connection definition, an array of connection coefficients can be constructed, each element of which shows the direct connection between a given pair of words.

2. *Defining clumps of index words*

Given an array of connection coefficients, we wish to identify subsets of words that are homogeneous by some criterion. One such criterion uses the notion of the *bias* of a word to a subset. Suppose that the total list of index words is arbitrarily divided into two groups (subsets), not necessarily of equal size. The bias of a word to one of these subsets is defined as the difference between its total connections to the words in the subset and its total connections to the words in the other subset. Given this measure we can specify a particular criterion of homogeneity that states that if the universe has been divided so that, for one of the subsets, all of its members have positive or zero bias, and all non-members have negative bias, than that subset is a clump. Needham has termed subgroups of this type "GR-clumps."

Although the concept of a GR-clump is a simple one, the process of locating clumps is quite complex. A procedure is described in an earlier paper (6), and has been implemented on a large-size digital computer.

3. *Interpretation of a clump*

In the context we have been describing, a clump is a loose group of words that possesses internal homogeneity, as measured by co-occurrence characteristics in a particular collection of documents. Put another way, membership of a word in a clump is evidence of a con-

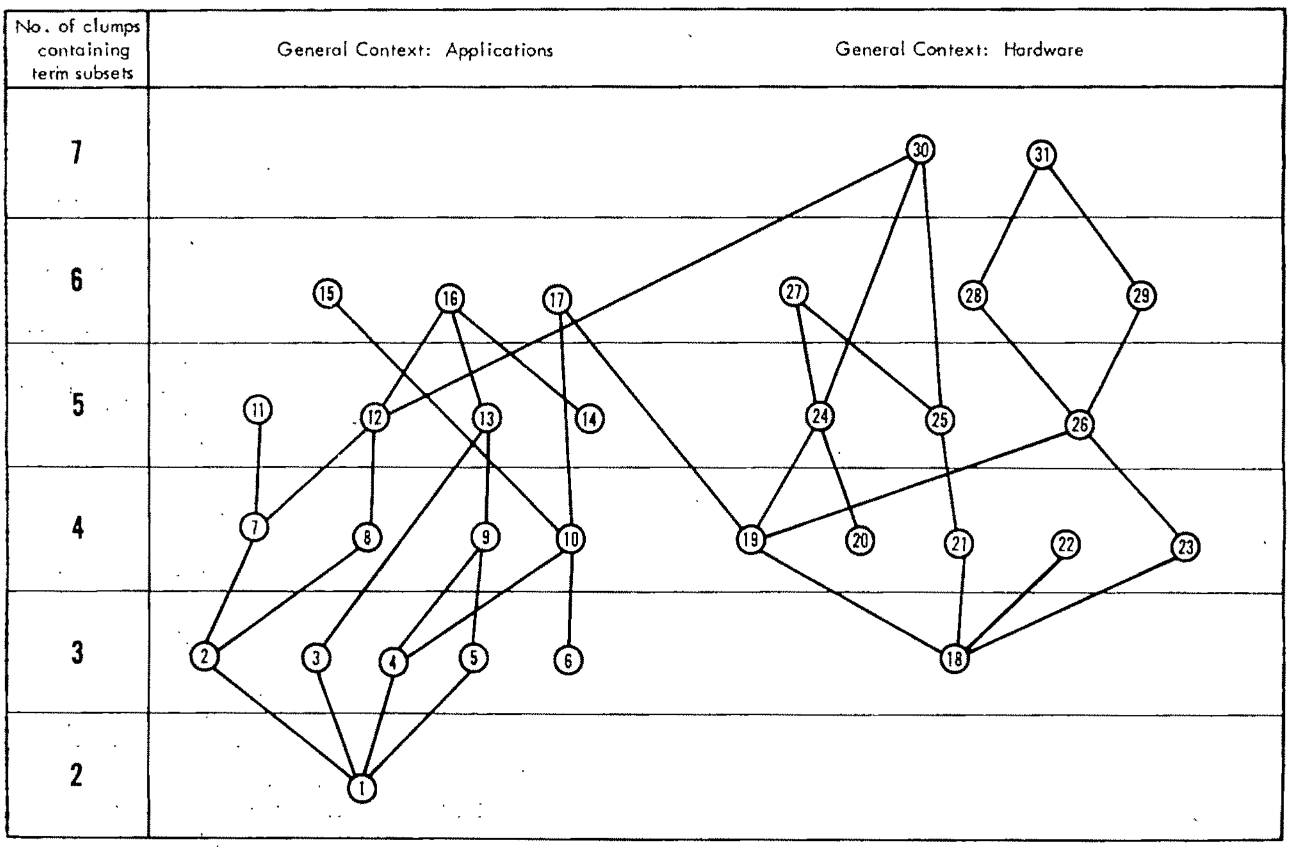


FIG. 1. Strong Term Associations Implicit in GR-Clump Structure, Connection Definition 2. (See Table 1 for Contents of Numbered Subsets)

textual relationship, within the collection, between the word and the other clump members.

Words may, of course, belong to more than one clump, and it is the information on the distribution of clump membership of words that forms the basis for an associative retrieval operation. That is, we may regard the membership and non-membership of a given word in a set of clumps as defining its properties of contextual distribution. This provides efficient and compact information on the pattern of index word associations in a particular collection.

• Clumping Experiments

Using the connection definitions described above, clumps of index words were found in a data set supplied by H. Borko of System Development Corporation. The data characterize the use of 90 index words in 260 document abstracts published in the March and June issues of the 1959 IRE *Transactions on Electronic Computers*, covering topics dealing with computing hardware and computer applications.

Using connection definition 1, 19 clumps were found, and using definition 2, eight clumps were located. The mean sizes of the two sets of clumps were 52 and 49 elements respectively.

The type of associations generated by a structure of clumps is shown in Figure 1, which diagrams only the strongest association paths implicit in the set of eight clumps found under connection definition 2. The circled numbers identify small groups of words (subsets) that have the same distribution of clump membership, i.e., similar contextual distributions in the collection (the subsets are identified in Table 1.) the numbers at the left of the chart indicate the number of clumps in which given subsets uniquely co-occur. For example, the words forming subset 31 co-occur in seven of the eight clumps found under connection definition 2. The heavy connecting lines show inclusion relations. Subset 31, for example, co-occurs with subsets 28 and 29 in six clumps.

It is possible to gain from this mapping some feeling for the complex pattern of word associations that are implicit in even a small clump structure. Quite interestingly, there is a partition of the association map into two distinct major topical areas of associations, with relatively weak cross-connections.

In summary, clumps of words are subsets that possess an internal homogeneity based on empirical evidence of co-occurrence in a particular document collection. If we regard the distribution of a word across the set of clumps as summarizing, in a certain defined sense, its observed associational characteristics, we can use this information as a basis for defining associations in the entire set of words used to describe a collection of documents, and as information for associative retrieval operations.

TABLE 1

Key to Numbered Term Subsets in Figure 1

Subset number	*Terms*
1	complexity, language, Uncol
2	arithmetic, expressions
3	mechanical, translation
4	bound, definition, parity
5	chess, mechanisms, process, program, programming, programs
6	pseudo-random, random
7	square
8	average, differential, division, equation, equations, multiplication solution, traffic
9	character, delays, Monte Carlo, shuttle, stage, unit
10	numbers
11	abacus, boolean, functions, matrix
12	diffusion, error
13	characters, office
14	section
15	simulation
16	analog, control, function, generator, plane
17	code, conversion, elements
18	added, carry, network, networks, scientific, synthesis
19	communications, register, decoder, shift, wire
20	circuit, circuits, counter, logic, pulse, transistor, transistors
21	storage
22	switching
23	fields
24	element
25	barium
26	file, information, library, magnetic, processing, tape
27	memory
28	transmission
29	printed, recording
30	side
31	coding, compressions, film, speech

• An Associative Retrieval Model

If a number of clumps of index words have been found, it is possible to set up two simple classification arrays that will form the search spaces for associative document retrieval operations. One array is a simple binary matrix that indicates whether or not a given word is a member of a given index word clump. The second array indicates, for each document, how many of its index words are contained in each of the index word clumps. This information is referred to as the "clump list" of a document.

A retrieval request will consist of a string of index words, and the retrieval problem is to produce a list of documents that are ordered on some scale of relevancy to the request.

Corresponding to any retrieval request will be a clump list, specifying the clumps to which the words contained

in the request belong. One possible definition of the prospective relevancy of a document to a given request is to require that the clump list of a document enclose the clump list of a request. This means that the contextual distribution of the document descriptors are at least coextensive with the request. It turns out, however, that this condition is over-restrictive, since it can lead to the exclusion of documents that possess some index words included in the request but do not satisfy the condition. Consequently, we define a relevant document to be one that either (a) contains index words included in the request, or (b) possesses a clump list that encloses the clump list of the request. Only such documents are considered for output, and the remainder of retrieval processing is concerned with computing relevance weights for the relevant set of documents. The weighting scheme takes into account the number of index word matches between request and document, the extent to which the contextual distribution of a document coincides with that of the request, and normalizes relevance scores to adjust for the number of index words used to describe a document. The associativity of the retrieval process is realized by the relevancy criterion, since it is possible for a document to possess a relevant clump list without having any explicit word matches with the input request. The model is described in more detail in a recent paper (7).

• Retrieval Experiments

The retrieval model has been programmed for use on a digital computer, and a number of retrieval experiments made within the collection of 260 documents, using the clump structures of connection definitions 1 and 2 noted above. One set of experiments was concerned with informal validation of the model; the results are summarized in Table 2. (Structures 1 and 2 refer to the clump structures of connection definitions 1 and 2, respectively.) The experimental technique was as follows:

(a) Four individuals with general familiarity of the subject fields covered by the set of 260 documents were given four randomly selected retrieval requests and asked to independently prepare lists of documents relevant to the requests by scanning the 260 abstracts and identifying documents on a three-valued relevancy scale ranging from most relevant (1) to possibly relevant (3).

(b) The manually prepared lists for a given request were consolidated and a sublist of documents most relevant to the request was prepared. This sublist comprised documents rated with a value of 1 by at least two of the four individuals, or rated with a value of 1 by one individual and rated 2 by at least two others.

(c) Comparisons of manual and automatic retrievals are given in Table 2.

Request 1 asked for documents dealing with language translation. Request 2 asked for documents dealing with circuitry in analog computers. Requests 3 was for documents on simulation. Request 4 called for documents dealing with programming languages.

Using the rule outlined in (b) above, 10 documents most relevant to the first request were identified, from a total set of 19 documents retrieved by the four investigators. The retrieval algorithm produced ordered lists of 104 documents using structure 1, and 105 documents using structure 2. In the upper fourth of the output list from structure 1, 7 of the 10 most relevant documents were located, and in the upper fourth of the structure 2 output list, 9 of the 10 most relevant documents. The algorithm failed to retrieve one of the most relevant documents using structure 1, but retrieved all the relevant documents using structure 2.

The table indicates the generally satisfactory performance of the retrieval model and confirms the reasonableness of the definition of relevance used.

Considering the overall recall and relevance (precision) ratios, it can be seen that recall ratios in the test requests were high (i.e., very few relevant documents were not retrieved); relevance ratios were characteristically smaller (of the order of 10 per cent). However, since the output lists are ordered, it is interesting to note that the relevance ratios are significantly much higher in the upper portions of the output lists (roughly between 25

TABLE 2

Request Number	Number of most relevant documents identified	Total number of documents retrieved: Manual	Automatic Structure 1	Automatic Structure 2	Number of most relevant documents in upper fourth of output lists: Structure 1	Structure 2	Number of most relevant documents in remainder of output lists: Structure 1	Structure 2	Number of most relevant documents not retrieved: Structure 1	Structure 2
1	10	19	104	105	7	9	2	1	1	0
2	14	63	89	100	10	10	4	4	0	0
3	15	43	119	94	8	11	6	1	1	3
4	12	32	115	181	8	11	3	1	1	0

Comparison of Manual and Automatic Retrievals

per cent and 50 per cent in the upper fourth of the output lists), and that recall ratios are still of the order of 50-70 per cent.

These results have suggested that if this type of retrieval process is implemented in a real-time computer operation in which the requester is in direct communication with the system, efficient retrieval in large collections can be attained.

For request 1, seven of the 10 most relevant documents identified occur in the first 26 documents of the 104 which were retrieved. If a requester were presented with these top 26 documents, he could indicate the seven most relevant. On the basis of this information, the system could reorder the remaining 78 documents on its output list to bring to the top those documents most likely to serve the needs of the requester.

• Summary

The limited experiments described in this paper indicate the potential utility of clumping techniques as a basis for large-scale associative document retrieval. This conclusion is based on several considerations arising from the work described, and from work currently in progress:

1. Clumping associations appear to provide a valid basis for retrieval.
2. The efficiency of retrieval, particularly as measured by the total number of documents retrieved in response to a request, will not be degraded in large collections, as is the case with coordinate systems.
3. Clumping techniques can be applied to compute associations in very large sets of data. We are currently experimenting with computing techniques capable of rapid identification of clumps in data sets of the order of 1,500 index words, and expect to have programs capable of handling much larger sets (of the order of 10,000 index words) later this year.

References

1. LITTLE, INC., ARTHUR D. 1963. *Centralization and Documentation:* Final Report to the National Science Foundation (Cambridge: July).
2. PARKER-RHODES, A. F., and NEEDHAM, R. M. 1960. *The Theory of Clumps* (Cambridge, England: CLRU).
3. NEEDHAM, R. M. 1961. *The Theory of Clumps* II (Cambridge, England: CLRU).
4. *Research on Information Retrieval Classification and Grouping,* 1957–61, (Cambridge, England: CLRU, 1961).
5. NEEDHAM, R. M. 1962. *A Method for Using Computers in Information Classification* (Presented at IFIPC, Munich, mimeo.).
6. DALE, A. G., and DALE N. 1964. *Some Clumping Experiments for Information Retrieval:* LRC 64–WPI1 (Austin: Linguistics Research Center, February).
7. DALE, A. G., and DALE, N. 1964. *Clumping Techniques and Associative Retrieval* (Presented at NBS–ADI Symposium on Statistical Association Methods for Mechanized Documentation, March).

An Application of a Type of Matrix to Analyze Citations of Scientific Papers

ERI YAGI *

Institute of Physics
College of General Education
University of Tokyo
Japan

• Introduction

A scientific paper has a series of citations belonging to the previous literature. In order to determine a linkage pattern of scientific papers through their citations, an analysis of a group, whose papers are frequently cited by themselves, is useful. It is possible to investigate the internal connections among papers in such a group by examining their citations. The use of a type of matrix for the analysis was suggested by Derek Price (1) while the techniques of the matrix analysis had been developed for the study of group structure in the fields of sociology and psychology.

The main purpose of this study is to investigate the measurements by which the internal connections of a group of scientific papers are measured.[1] The techniques of using the type of matrix, suggested by Derek Price, will be further developed to carry out the analysis of a certain type of group of scientific papers. The collected papers by members of the Japanese Group of Nuclear Physics will be used as an example. It has been suggested that these collected papers must have close internal connections among themselves because of the character of the Japanese Group of Nuclear Physics as a scientific school.

In Part I of this paper the type of matrix to analyze citations of scientific papers will be shown in general.

$$X = \begin{array}{cc} & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & \cdot & \cdot & \cdot & \cdot & n \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ \cdot \\ \cdot \\ \cdot \\ n \end{array} & \begin{bmatrix} 0 & x12 & x13 & x14 & \cdot & \cdot & \cdot & \cdot & x1n \\ 0 & 0 & x23 & x24 & \cdot & \cdot & \cdot & \cdot & x2n \\ 0 & 0 & 0 & x34 & \cdot & \cdot & \cdot & \cdot & x3n \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & x4n \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot \\ & & & & & & & & \\ & & & & & & & & \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix} \end{array}$$

FIG. 1. Matrix of One-link Connections among n Scientific Papers. The Rows Represent the Papers Cited, and the Columns the Source Papers.

* United States Public Health Service Training Grant, No. 2G–450, at the Department of History of Science and Medicine, Yale University, 1960-63.

[1] For general discussions on the measurement on science, see my paper on the subject: Eri Yagi, *Scientific Papers of the College of General Education*, University of Tokyo, 14, 129, 1964.

[2] The Japanese Group of Nuclear Physics was originated by Yoshio Nishina in the 1930's and has been developed by Hideki Yukawa, Shin-itiro Tomonaga, Mitsuo Taketani, Shoichi Sakata, and other Japanese physicists.

In Part II of this paper several new types of measurements will be proposed to indicate the internal connections of a group of scientific papers by applying the above type of matrix to the collected papers of the Japanese Group of Nuclear Physics.

Part I. A Type of Matrix to Analyze Citations of Scientific Papers

The construction of a type of matrix to analyze citations of scientific papers in a group will be shown in this section. The matrix is constructed by listing the papers in chronological order both along the rows and down the columns. The rows represent the papers cited and the columns the source papers.

Let Xij be called an element of the matrix, Xij being the element in the ith row and jth column. The row-suffix i ranges over the values, 1,2,...., n, and the column-suffix j over the same values, 1,2,...., n. Here n is the total number of papers in the group. The Matrix as a whole will be denoted by X.

In this matrix X, one (1) means a citation and zero (0) no citation so that if the element $X_{12} = 1$, this means that paper No. 2 cites paper No. 1; if $X_{12} = 0$, paper No. 2 does not cite paper No. 1. This is called a matrix of one-link connections, which indicates direct links among n scientific papers. Figure 1 is the matrix of one-link connections. In this matrix X, the entire half of elements below the main diagonal is zero because a paper cites only papers which have been previously published; the elements on the main diagonal line are also zero because individual papers do not cite themselves.

Multiplying the original matrix, X, by itself, one can determine how papers in the group are indirectly linked. This squared matrix, X^2, is called a matrix of two link connections, which indicates two-step indirect links among n scientific papers. The elements in the ith row and jth column of X^2 is generally shown as $\sum_{k=1}^{n} X_{ik}X_{kj}$. Figure 2 is the matrix X^2 of two-link connections. In this matrix, the elements on the diagonal line next to the main diagonal are also zero.

The element in the first row and third column of X^2 is equal to $X_{12}X_{28}$[3]. If $X_{12}X_{28} = 1$, paper No. 1 is said to have a two-link connection with paper No. 3, i.e. paper

[3] $\sum_{k=1}^{n} X_{1k}X_{k3} = X_{11}X_{13} + X_{12}X_{23} + X_{13}X_{33} + :::: + X_{1n}X_{n3}$
$= X_{12}X_{23}$, because all other products are zero.

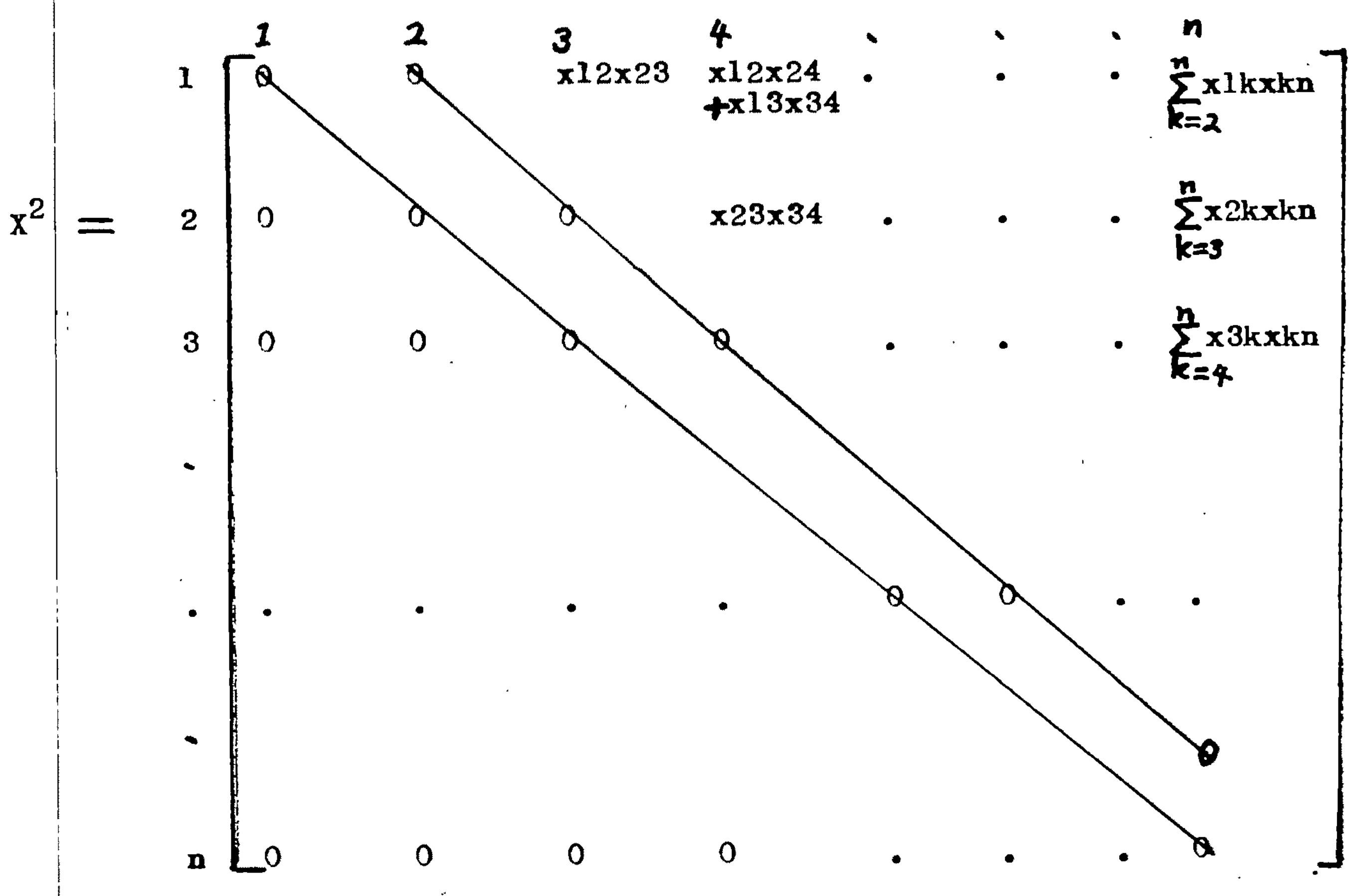


Fig. 2. Matrix of Two-link Connections among n Scientific Papers.

No. 3 cites paper No. 2, and paper No. 2 cites paper No. 1. This will be denoted by 1→2→3. (Note that paper No. 3 need not cite paper No. 1 directly). If $X_{12}X_{23} = 0$, paper No. 1 is said to have no two-link connection with paper No. 3.

The element in the first row and fourth column of X^2 is equal to $X_{12}X_{24} + X_{13}X_{34}$. If $X_{12}X_{24} + X_{13}X_{34} = 2$, there are two two-link connections between paper No. 1 and paper No. 4, i.e., both 1→2→4, and 1→3→4. If $X_{12}X_{24} + X_{13}X_{34} = 1$, there is one two-link connection between paper No. 1 and paper No. 4, i.e., either 1→2→4, or 1→3→4. If $X_{12}X_{24} + X_{13}X_{34} = 0$, there is no two-link connection between paper No. 1 and paper No. 4.

It is apparent that the original matrix, X, may also be raised to higher powers to obtain the three-link, four-link or even more link connections among n scientific papers in the group.

$$\sum_{k=1}^{n} X_{1k}X_{k4} = X_{11}X_{14} + X_{12}X_{24} + X_{13}X_{34} + \cdots + X_{1n}X_{n4}$$
$$= X_{12}X_{24} + X_{13}X_{34}, \text{ because all other products are zero.}$$

Part II. An Application of the Type of Matrix to the Analysis of a Group of Collected Papers by Members of the Japanese Group of Nuclear Physics

The collected papers of the Japanese Group of Nuclear Physics will be analyzed as an example in this section. This group of papers on the meson theory is collected in the *Supplement of the Progress of Theoretical Physics*, No. 1, in 1955 in commemoration of the 20th anniversary of the proposal of the theory by Hideki Yukawa (3). For this proposal Yukawa won the Nobel Prize for Physics in 1949. The group consists of 34 important papers on the meson theory by members of the Japanese Group of Nuclear Physics in the period of 1935-50. This group of publications will be called Is in the following discussion.

1. Measurements concerning papers inside of group Is.

Figure 3 shows the matrix of one-link connections among 34 papers which belong to group Is. These papers are listed along the rows and down the columns by the method which has been given in the previous section. This matrix in Fig. 3 is designated as MIs.

There is a triangular area in the right part of MIs where all the elements are zero. This indicates the time dependence of citations among scientific papers: in later papers, earlier papers are not cited.

The sum of elements of MIs is 67, by which one can indicate the strength of internal connections of group Is. In order to compare different groups which do not consist of the same number of papers, it may be useful to take the ratio between the obtained sum of elements, 67, and the maximum possible sum of elements for the special

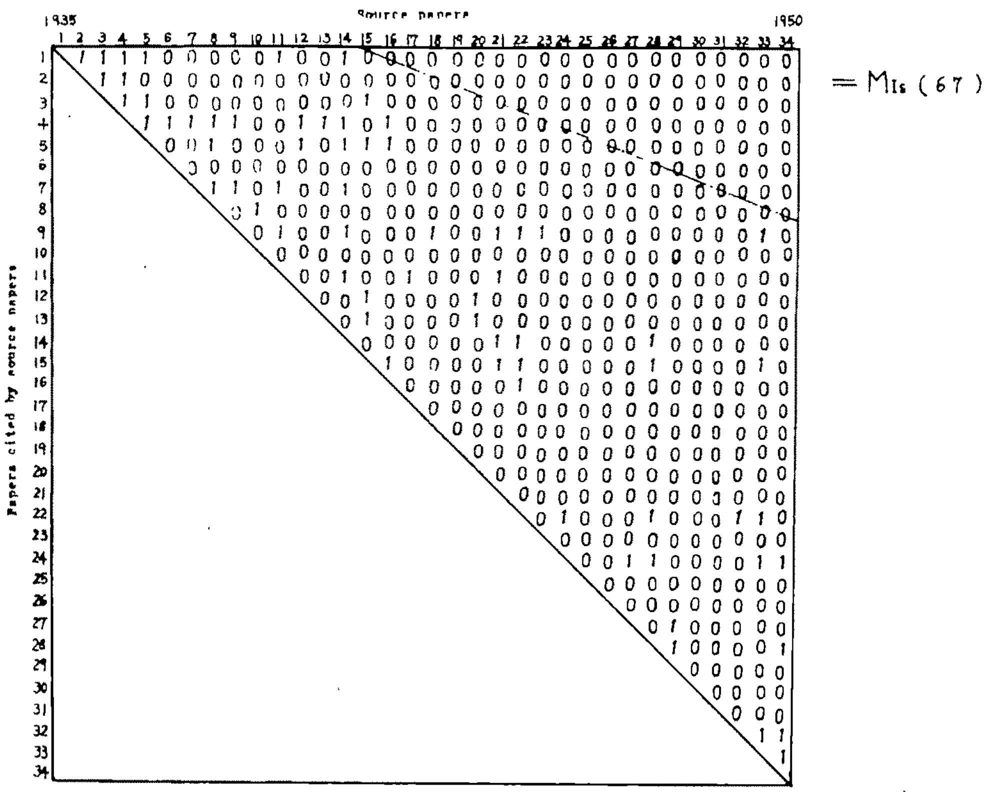


Fig. 3. Matric of the One-link Connections between Papers by Members of the Japanese Group of Nuclear Physics in the *Supplement of the Progress of Theoretical Physics*, No. 1.

case of the matrix in which each paper is directly connected with all the papers previously published. The maximum sum is 561 for 34 papers because the maximum sum is ½ n(n — 1) for n scientific papers.

In order to construct the matrix of two-link connections among 34 papers which belong to group Is, matrix MIs is multiplied by itself. This squared matrix in Fig. 4 is $(\mathrm{MIs})^2$.

The sum of elements of $(\mathrm{MIs})^2$ is 157, by which one can also indicate the strength of internal connections of group Is. In order to compare different groups which do not consist of the same number of papers, it may be useful to take the ratio between the obtained sum of elements, 157, and the maximum possible sum of elements for the special case of the squared matrix in which each paper has two-step indirect connections (e.g. 1→2→3) with all the papers previous published. The maximum sum is 5,984 for 34 papers because the sum is 1/6 n(n — 1) (n — 2) for n scientific papers.

2. Measurements concerning papers outside of group Is.

Here new measurements to indicate the strength of internal connections of group Is will be proposed by the use of papers which are cited by group Is, but do not belong to group Is.

Table 1 is the distribution of citations according to the number of times that each paper is cited by group Is. The first line of the table, for example, is read as follows: One paper, which is outside of group Is, is cited 8 times by papers which are inside of group Is. Therefore, there are 8 such citations.

For convenience, one can divide these papers outside of group Is into the following two groups, i.e., group Ii and group Iw. Group Ii consists of 34 papers cited frequently by group Is, but not belonging to group Is: thirty of these papers are cited more than twice, and four of those are cited twice. Group Iw consists of 34 papers cited less frequently by group Is: These 34 papers are a random sample of papers, cited once by group Is, but belonging to neither group Is nor group Ii. Both groups Ii and Iw consist of papers written by foreign physicists not by Japanese.

Figures 5 and 6 are the matrix of one-link connections among papers of group Ii, and that of group Iw, respectively. Then, the matrix in Fig. 5 will be called MIi, the matrix in Fig. 6 MIw.

The sum of elements of MIi is 54, and that of elements of MIw is 11, while that of elements of MIs is 67 as mentioned. This means that papers which belong to group Is have the strongest one-link connections among these three groups, Is, Ii, and Iw. On the other hand, papers which belong to group Iw have the weakest one-link connections.

By the matrix multiplication, the squared matrix of MIi, $(\mathrm{MIi})^2$, and the squared matrix of MIw, $(\mathrm{MIw})^2$

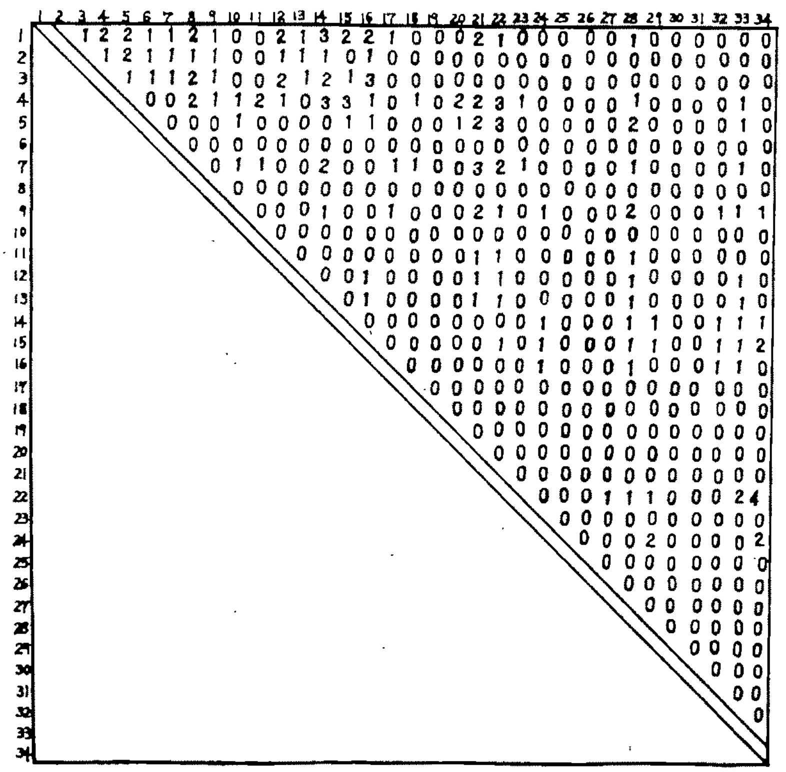


FIG. 4. Matrix of the Two-link Connections among Papers by Members of the Japanese Group of Nuclear Physics, Group Is.

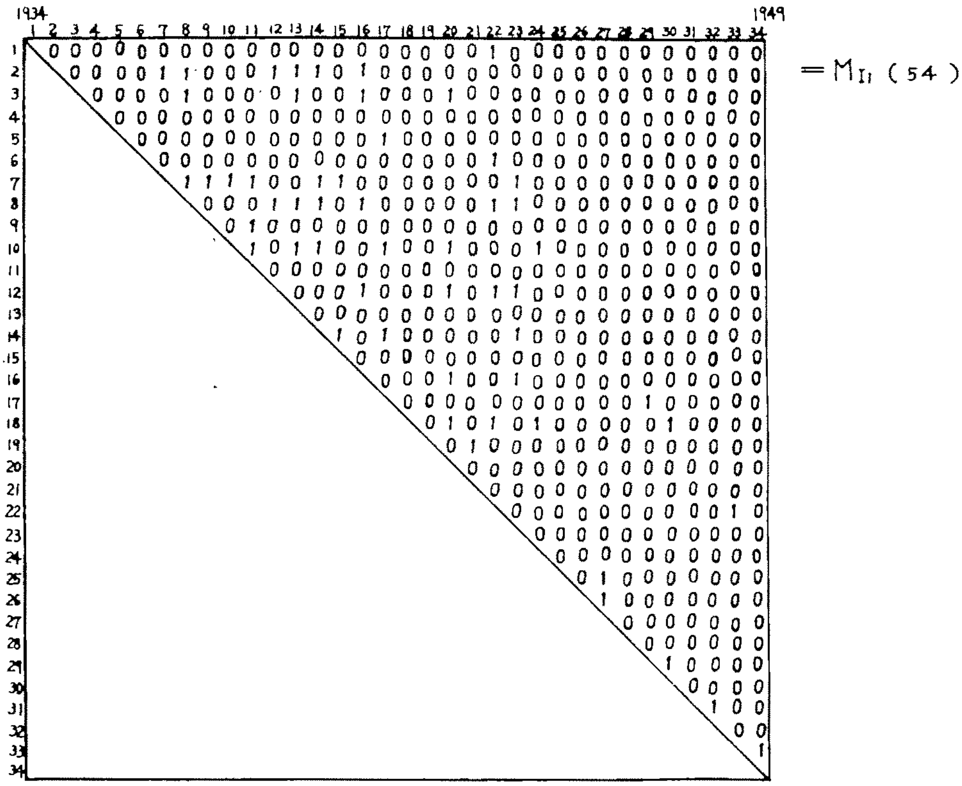


FIG. 5. Matrix of the One-link Connections between Papers of Group Ii.

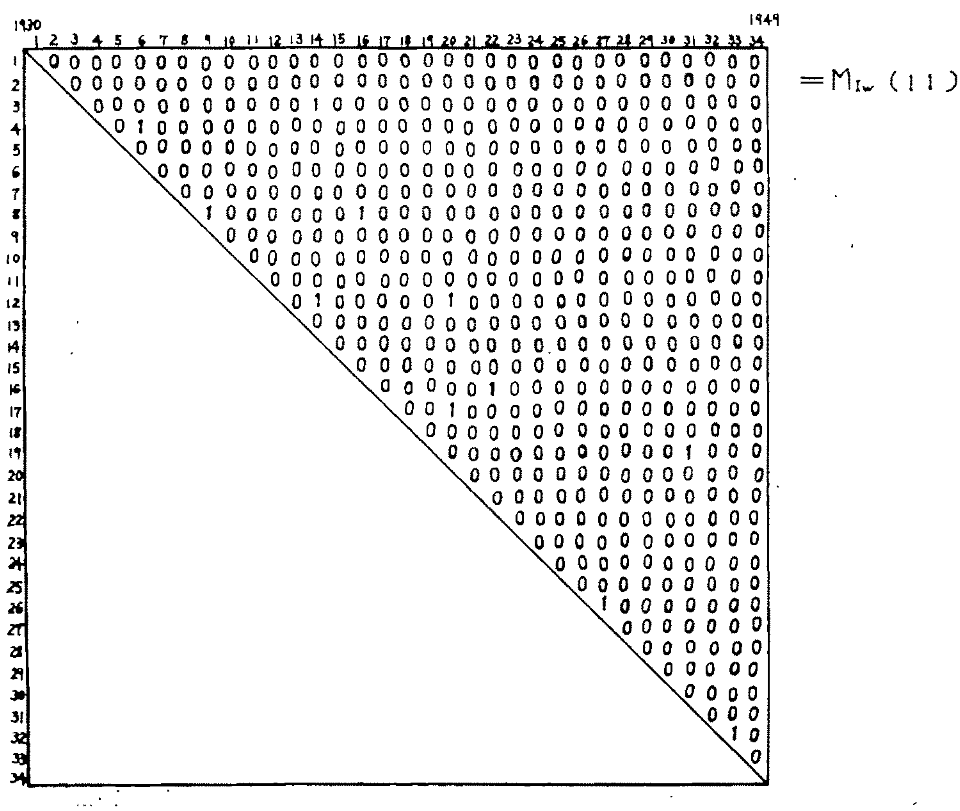


FIG. 6. Matrix of the One-link Connections between Papers of Group Iw.

are made. (See Fig. 7) The sums of elements of these squared matrices are as follows:

Sum of elements of $(MIi)^2 = 70$
" " " " $(MIw)^2 = 0$
while that of " " $(MIs)^2 = 157.$

These sums show that papers which belong to group Is have the strongest two-link connections among these three groups, Is, Ii, and Iw, and that papers which belong to group Iw have the weakest two-link connections. This order of the strength of two-link connections is the same order as that of one-link connections.

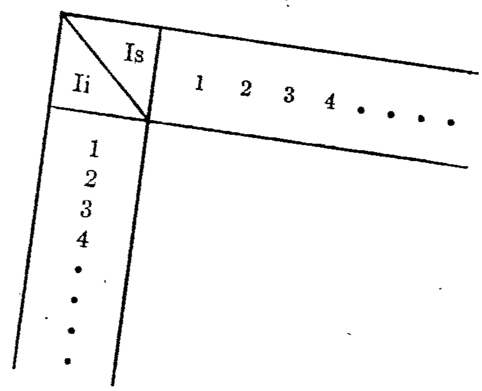
In addition to the values of the sum of elements of MIs and $(MIs)^2$, one may use the values of the sum of elements of MIi, $(MIi)^2$, MIw, and $(MIm)^2$ as measurements to indicate the strength of internal connections of group Is because these two groups Ii and Iw are brought out in connection with group Is. The order of these three groups, Is, Ii, and Iw, in the strength of one-link, or two-link connections may be used as another measurement.

3. Measurements concerning groups Is, Ii, and Iw.

In the first place, the construction of the matrix of one-link connections between every two of groups Is, Ii, and Iw will be given to indicate the strength of mutual connections.

In matrix MIis, 34 papers, belonging to group Is, are listed along the rows, and 34 papers, belonging to group Ii, down the columns in chronological order. The row corresponds to the paper being cited, and the column corresponds to the paper which cites it.

MIis =



MIsi, MIwi, MIiw, MIws, and MIsw are constructed in a similar fashion. See Figs. 8, 9, 10, 11, 12 and 13. The element in the first row and first column of MIis is one (1), this means that paper No. 1 of group Is cites paper No. 1 of group Ii. The element in the second row and first column of MIis is zero (0), this means that paper No. 1 of group Is does not cite paper No. 2 of group Ii.

By adding the elements of each matrix, the following sums are obtained:

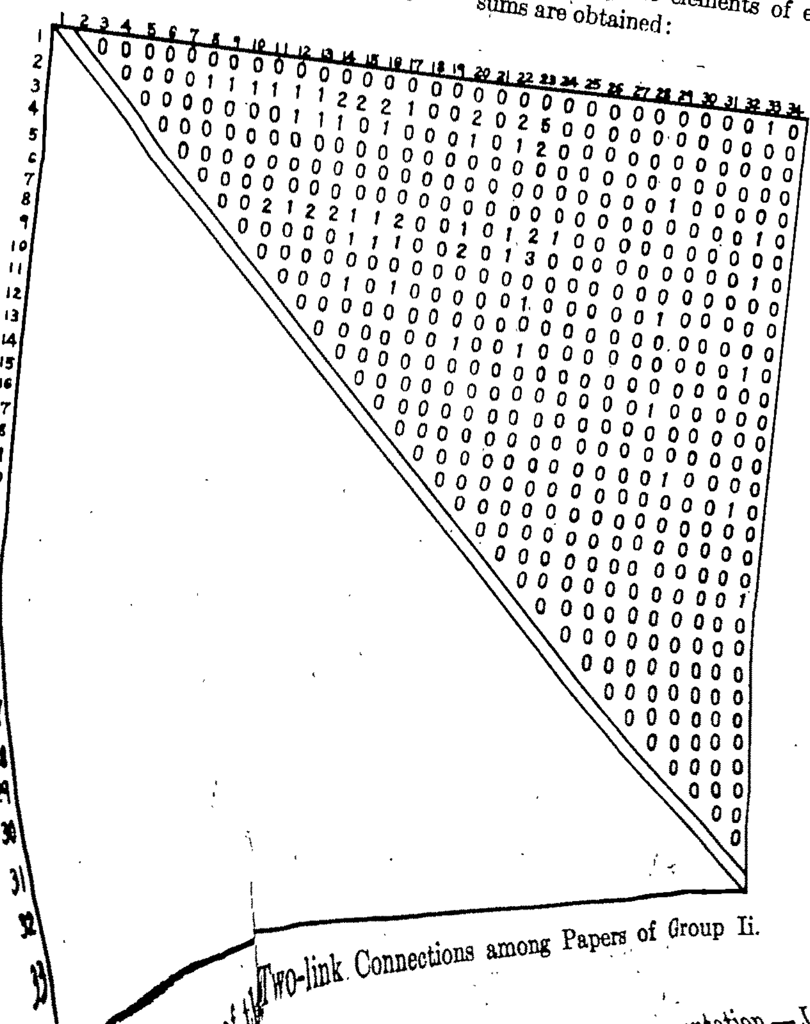
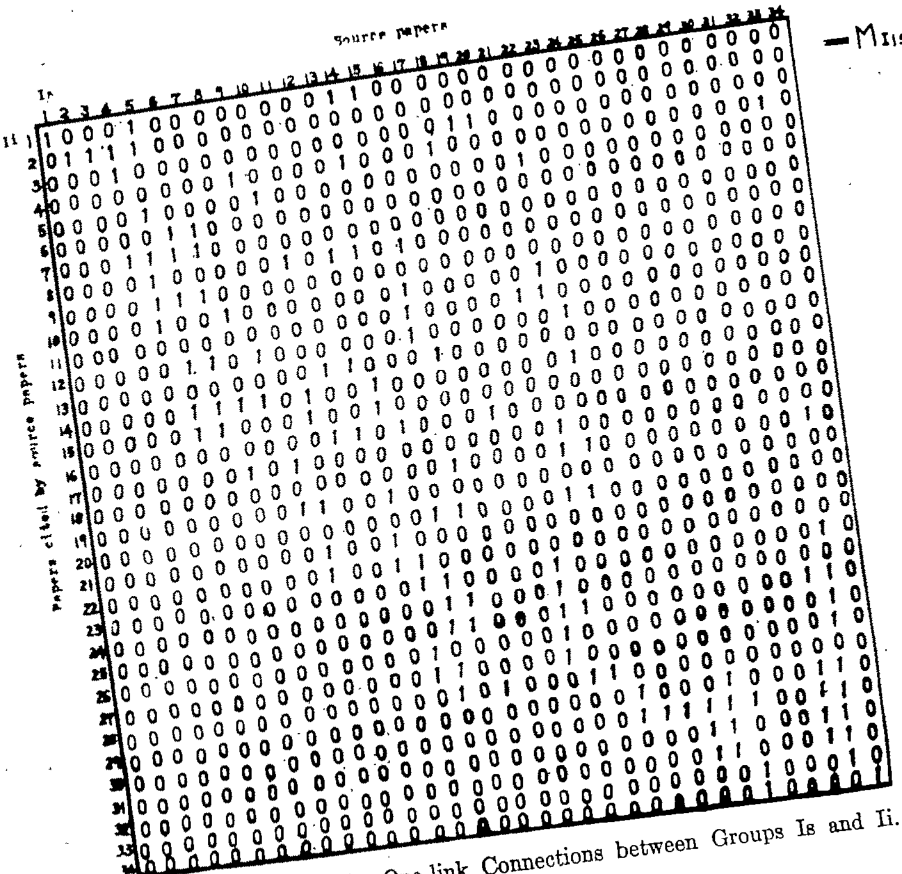
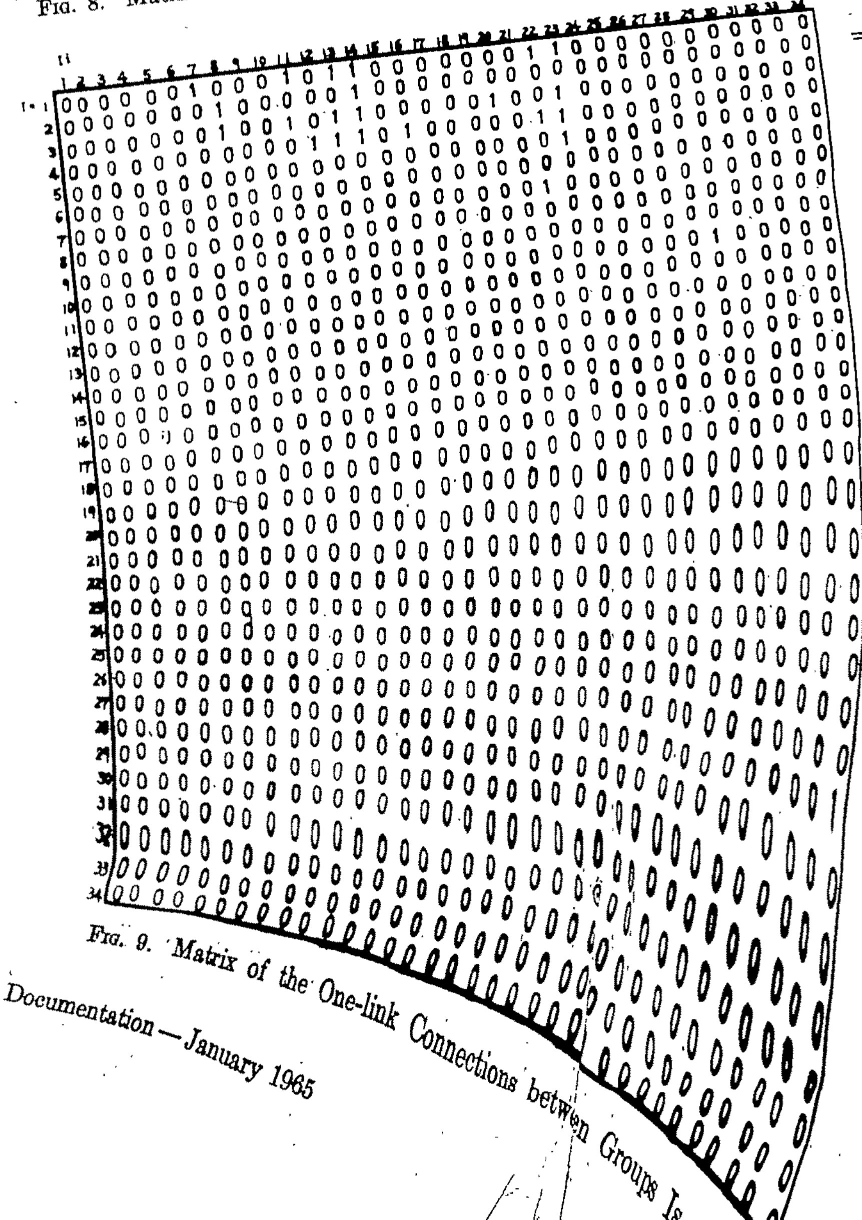


Fig. 7. Matrix of the Two-link Connections among Papers of Group Ii.



— $M_{I_iI_s}$ (129)

Fig. 8. Matrix of the One-link Connections between Groups Is and Ii.



= $M_{I_sI_i}$ (24)

Fig. 9. Matrix of the One-link Connections between Groups Is and Ii.

Ii

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Iw 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

$= M_{Iwi}$ (16)

Fig. 10. Matrix of the One-link Connections between Groups Ii and Iw.

Iw

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Ii 1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

— M_{Iiw} (32)

Fig. 11. Matrix of the One-link Connections between Groups Ii and Iw.

Is

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Iw 1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

$= M_{Iws}$ (34)

FIG. 12. Matrix of the One-link Connections between Groups Iw and Is.

Iw

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Is 1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$= M_{Isw}$ (13)

FIG. 13. Matrix of the One-link Connections between Groups Iw and Is.

Sum of elements of MIis $= 129$

Sum of elements of MIis	=	129
MIsi	=	24
MIwi	=	16
MIiw	=	32
MIws	=	34
MIsw	=	13

Therefore, the sum of elements of MIis is larger than that of MIsi. This means that group Ii is cited by group Is with greater frequency than group Is is cited by group Ii. The sum of elements of MIiw is larger than that of MIwi. This means that group Ii is cited by group Iw with greater frequency than group Iw is cited by group Ii. The sum of elements of MIiw is larger than that of MIwi. This means that group Iw is cited by group Is with greater frequency than group Iw is cited by group Ii. values of the sum of the above 6 matrices are arranged as follows: Sum of elements of MIis $>$ MIws $>$ MIiw $>$ MIsi $>$ MIwi $>$MIsw.

The above analysis suggests the characteristics of group Is that papers inside of group Is are closely connected with themselves but not completely closed within themselves. It has been shown that papers inside of group Is are also connected with other papers outside of group Is through the citations.

• Summary and Conclusions

In the previous sections the group of papers of the Japanese Group of Nuclear Physics, which is called group Is, has been analyzed as an example by the matrix analysis.

It is possible for group Is to indicate the internal connections by the following measurements:

1. The matrix of one-link connections of group Is; this matrix is called MIs. The value of the sum of elements of MIs.
2. The matrix of two-link connections of MIs; this matrix is $(\mathrm{MIs})^2$. The value of the sum of elements of $(\mathrm{MIs})^2$.
3. The matrix of one-link connections of papers which are frequently cited by group Is, but not belonging to group Is. This group of papers, which is outside of group Is, is called group Ii; the matrix of group Ii is called MIi. The value of the sume of elements of MIi.
4. The matrix of two-link connections of MIi; this matrix is $(\mathrm{MIi})^2$. The value of the sum of elements of $(\mathrm{MIi})^2$.
5. The matrix of one-link connections of papers which are less frequently cited by group Is, but belonging to neither group Is nor group Ii. This group of papers, which is outside of group Is, is called group Iw; the matrix of group Iw is called MIw. The value of the sum of elements of MIw.
6. The matrix of two-link connections of MIw; this matrix is $(\mathrm{MIw})^2$. The value of the sum of elements of $(\mathrm{MIw})^2$.

It is possible to indicate the mutual connections between every two of groups Is, Ii, and Iw by matrices MIis, MIsi, MIiw, MIwi, MIws and MIsw. MIis, for example, are constructed by listing papers of group Is along the rows, and papers of group Ii down the columns in chronological order.

Group Is has the characteristics that papers inside of group Is are closely connected with themselves but not completely closed within themselves. This has been shown by the existence of groups Ii and Iw, which are outside of group Is.

It is hoped that other groups of scientific papers are analyzed by the method that has been developed in this study.

• Acknowledgments

The research on which this paper is based was carried out at the Department of History of Science and Medicine, Yale University, under the direction of Professor Derek Price, to whom I am indebted for much encouragement and criticism. Thanks are also due to Dr. George Sorger of the Department of Microbiology for kind assistance in improving the understandability and clarity of the manuscript.

References

1. This was suggested at the seminar on Sociology and Politics of Recent Science, Department of the History of Science and Medicine, Yale University, February 1963. See also Derek Price, "Why Does Science Cumulate", *Frontiers of Science and Philosophy*, Vol. 2, University of Pittsburgh Press (in press).
2. GLANZER, M., and GLASER, R. 1959. *Psychological Bulletin*, Vol. **56**, No. 5, Sept., 317. FESTINGER, L., SCHACHTER, S., and BACK, K. 1955. "Matrix Analysis of Group Structures", LAZARSFELD, PAUL F., and ROSENBERG, M., ed., *The Language of Social Research.* pp. 358–367.
3. *Supplement of the Progress of Theoretical Physics*, Kyoto, No. 1, 1955.

TABLE 1. Distribution of papers outside of Group Is according to the frequency with which they are cited by Group Is

No. of times each paper is cited (x).	*No. of papers (y)*	*No. of such citations (x y)*
8	1	8
7	1	7
6	1	6
5	2	10
4	15	60
3	10 + 2[a]	36
2	49 + 2[a]	102
1	194 + 34[a]	228
Total 36	273 + 38[a]	457

[a] Indicates numbers of papers by Japanese Physicists.

A Quantitative Measure of User Circulation Requirements and Its Possible Effect on Stack Thinning and Multiple Copy Determination

The strategy outlined in this article describes a possible aid to the librarian for thinning a library's stacks according to the criterion of user needs. The method uses the last circulation date as a parameter of user circulation requirements. The resulting stack collection would, by design, satisfy over 99 percent of the user circulation requirements and yet be of minimum size. Preliminary, but not yet fully validated research, indicates that the number of volumes in a library's holdings may be reduced by 60 to 70 percent and yet satisfy well over 99 percent of the user requirements. It may also be possible to determine which books should have multiple copies in order that user disappointment can be minimized. The effective result of this approach is that there now may be a quantitative method of maintaining the library's holdings at a reasonable level and, in addition, a reduction in the disappointment of the user who is unable to find the book he wants because of use by another.

RICHARD W. TRUESWELL

Associate Professor in Industrial Engineering
University of Massachusetts
Amherst, Massachusetts

• Stack Thinning and Weeding

Certain preliminary comments must be made by the author about stack thinning and weeding in order that this article may be considered in its proper perspective. There have been many arguments and debates about the merits of stack thinning or weeding. The problem seems to center itself about the situation of a user needing the book or books that have been removed and possibly discarded from the library's population of holdings. Conversely, if all books are held, even when the chance of use is remote, there is the problem of increased size of the library. Coupled with this is the fact that it is impossible to predict accurately the requirements of scholars and researchers delving into important but yet sometimes obscure areas of research. Also, if a library is to be truly great library, then it must maintain a complete collection, whether or not the volumes are used.

It should be emphasized, therefore, that the author does not advocate the arbitrary thinning and discarding of books from a library, even by the method described in this article, until a full evaluation and determination has been made by the administrators of the given library with respect to its overall goals and user responsibilities. The author is describing a technique that has possible value as an aid to stack thinning and multiple copy determination; however, the decision to thin or weed in a library and the subsequent storage or discarding of the weeded volumes is a policy decision. It must be made by the administration of the library, based on the overall objectives of the given library and its users.

It is felt that the information presented in this article may be of value to librarians for improving service to the user by increasing the probability of the user finding what he or she wants. This can be done by analytically predicting multiple copy needs. It may also be possible to optimize the size of the library holdings, with respect to user requirements, for those libraries where operating policy and objectives permit such an approach. The university library, for example, may be forced by its very nature as a research library to maintain all of its holdings but to separate them into a low-use storage area and a high-use holding area.. The latter could be used to satisfy a predefined, say 99%, proportion of user requirements. The low-use area would have limited access and the high-use area could be of the open stack variety. Again it is emphasized that these decisions are policy decisions and are not directly related to the methods described in this article.

• Introduction

One of the major problems confronting the librarian of today is how to cope with the increasing size of his

library's holdings. The continuing stream (see Meier, 1963) of incoming books must be offset by a comparable outflow of books that are no longer needed by the library, or the library will increase its holdings at a rate equal to the difference of these two flows. Many libraries are today constructing additions, extensions, and in some cases, entirely new libraries to cope with this and other problems relating to space. In other cases, librarians are attempting to hold the line by thinning stacks of those books no longer of use to the library and its users. The problem is somewhat complicated by the fact that serials and periodicals represent a different problem than monographs. By their very nature, most periodicals and some serials must be maintained over a considerable period of time, and little stack thinning can be accomplished except in the case of older periodicals put into some form of limited access storage. This article is primarily concerned with monographs, although the same approach could be applied to periodicals.

The usual approach to stack thinning is to have a professional staff member examine books in the stacks and to make a decision based on recorded use of the book, the age, the subject area, the title, the author, and many other factors, both qualitative and quantitative, with the end result lying somewhat in the subjective decision area. (see McGraw) This end result is sometimes of question-

TABLE 1

Deering Library Circulation
Previous Charge Date

Time period in months prior to circulation date of sample.	*% of sample not previously charged out during the cumulative time period.*
0	100%
1	89
2	76
3	68
4	59
5	51
6	49
7	42
8	39
9	38
10	35
11	32
12	29
18	24
24	17
36	11
48	8.2
60	5.8
Prior to 60th month	0

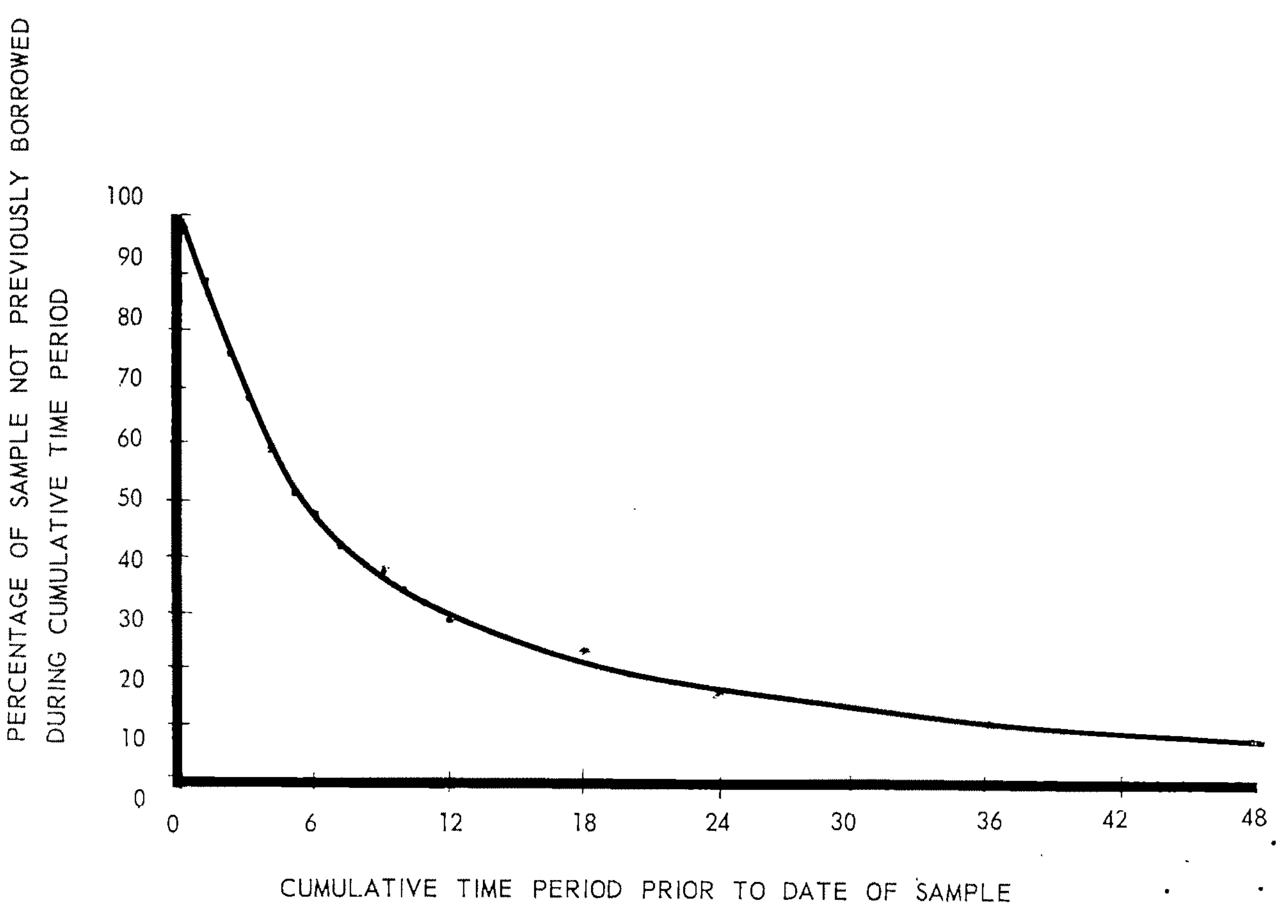


FIG. 1. Last Circulation Date Analysis (Deering Library).

able value because the librarian is not always sufficiently familiar with the subject field to make a correct determination of the value of the book for retention in the library. Therefore, some books are kept that are not really needed, and some books that should not be removed are discarded or taken to a remote storage area. The extent of these errors in thinning is not always realized because the user does not always indicate to the librarian that he was unable to find a particular book. It is suggested that the criteria for stack thinning should be designed to help the library satisfy the requirements of the users of the library. The entire approach of the method described in this article centers around this basic precept. It is realized that there are other reasons for keeping certain rarely, if ever used monographs. The extent of such rarely used and special collection items is, as previously stated, a major policy decision for the given library and will not be considered here.

It is felt that the primary purpose of a library is to supply the information and, in particular, the books required by the user when the user wants them. Several studies (Roy, 1963; Meier, 1963; Gnoza, 1951; Trueswell, 1964) show that there is approximately a fifty-fifty chance (actual probability found was .40 to .45) of the user *not* finding the book he wants in the stacks, given that the book is part of the library's holdings. Such a disappointment rate is rather large for a system designed to provide books for its users. This second problem can be approached by providing duplicate copies. However, the problem then becomes one of determining which books should have the duplicate copies and how many copies are necessary. There is also the question of whether or not there should be a non-circulating, core collection. The approach in this article also offers possible solutions to this problem.

Kilgour (1961), and later Trueswell (1964), approached the description of user circulation requirements through the statistic of age of the volume. This statistic, when translated into the percentage of holdings satisfying a given percentage of circulation, does not lend itself to thinning of monographs. It does, however, have possibilities for journals. Later studies by Kilgour (1962, 1964) show high use of certain medical journals in the Yale Medical Library. For example, 37 of the 1437 journals received furnished 49% of the recorded use.

• Basic Approach

As an approach to the problems described above, the last circulation date for books currently being charged was considered to be indicative of the use of the book. Table 1 is a cumulative distribution of this statistic for a sample taken from the circulation of the Deering Library at Northwestern University. It is generally accepted that certain books are used more frequently than others, which would lead one to expect that in any day's circulation there will be more books that have been previously charged within the last few months and fewer books that have been previously charged one or more years prior to this current circulation date. Thus we find that of the current circulation sample, 100% minus 89%, or in effect, 11% of the sample were books that were extremely popular and had been charged out at least once during the previous month. Similarly, 41% of the current sample had been charged out at least once during the previous four months' period. Figure 1 is a plot of the same information and is, in effect, a distribution of the current circulation sample with respect to the previously borrowed date for each book. It has been suggested (Trueswell, 1964) that this information could be used to predict a relatively low-cost method for conversion of the library's manual circulation control system to a punched card or computerized circulation control system. Again, as an illustration, approximately 30% of the current circulation sample is made up of books that have not been previously charged during the past twelve months. Similarly, about 8% had not been charged in the past 48 months. Thus, if the statistic of previous or past charge date is known, one can predict what portion of the circulation sample the book will fall into. Table 2 and figure 2 represent similar data for the Technological Institute of Northwestern University. Here we find that only 3% of the circulation sample had their last charge date occur sometime prior to the last 36 months.

These statistics can be expressed in another way; namely, to say that in the case of the Tech Library, 97% of the current circulation is made up of books that have been charged one or more times during the previous 36 months. Thus, if the holdings of the library were redefined to contain only those books that had been previously charged during the last 36 months, one could then expect to satisfy 97% of the current circulation requirements (giving rise to requests for the other 3% of circulation requirements). The actual distribution of the samples taken at the Technological Institute Library and the Deering Library went back further than 36 months and a less-than-1% level could be determined for each of these libraries. In the case of the Technological Institute Library, the over-99% figure was reached approximately after an eight year period. That is to say, over 99% of the current circulation sample at the Technological Institute Library was made up of books that had circulated at least once during the previous eight year period. Similarly for the Deering Library, it was found that the over-99% point was represented by books that had circulated at least once during the past twenty years. Thus, we might conclude that the sample of current circulation indicates current circulation to be made up of books (in the case of the Tech. Library) that circulated at least once during the previous eight year period. We now have a way to remove books from the stacks by using the following decision rule: Remove all books that have not circulated during the previous eight year period. After doing this, it

would be expected that no more than 1% of the users would be unable to find the books they required. Actually, this 1% level can be set at ¼ to ½% to give a more useful figure.

TABLE 2

Tech Library Circulation
Previous Charge Date

Time period in months prior to circulation date of sample.	*% of sample not previously charged out during the cumulative time period.*
0	100%
1	79
2	49
3	42
4	34
5	26
6	25
7	23
8	19
9	17
10	16
11	13
12	12
13	12
14	11
15	11
16	10
17	10
18	9
19	8
20	8
21	7
22	7
23	6
24	6
36	3
Prior to 36th month	0

The question that next presents itself is, if such a decision rule is put into effect, how many books will be removed from the holding of the library and thus what will be the size of the resulting optimal core collection? This figure can be predicted by again examining figures 1 and 2. If we use the eight year period for the Technological Institute which provides for 99% satisfaction of user requirements, we can generate or build a model collection of books based on the circulation that has occurred during the eight year period. If we make a rather loose assumption that the circulation pattern obtained from the current circulation sample is roughly equivalent to the circulation patterns for each of the past eight years, we can then predict the size of the core collection. For example, if we start eight years back in time, during the first month of circulation, 100% of the books circulated would not have previously circulated during this eight year period. Similarly, after six months of circulation, approximately 25% of these books (current circulation) would not have previously circulated. Thus, for each month in time during this eight year period, we have a percentage that indicates the percentage of books that have not previously circulated during this eight year period; or in effect, these are books that are now entering the population of books made up of books that have circulated one or more times during the eight year period. Thus, as each month goes by, we find a certain quantity of books that have not previously circulated during the time since the beginning of the point in time eight years past. We can therefore, using circulation figures over this time period, estimate the size of this collection by multiplying the percentage of the sample that has not previously circulated by the circulation for each successive month of the eight year period. As time progresses, in fact after 36 months, we find that only about 3% of the books being circulated have not previously circulated since the beginning of the eight year period.

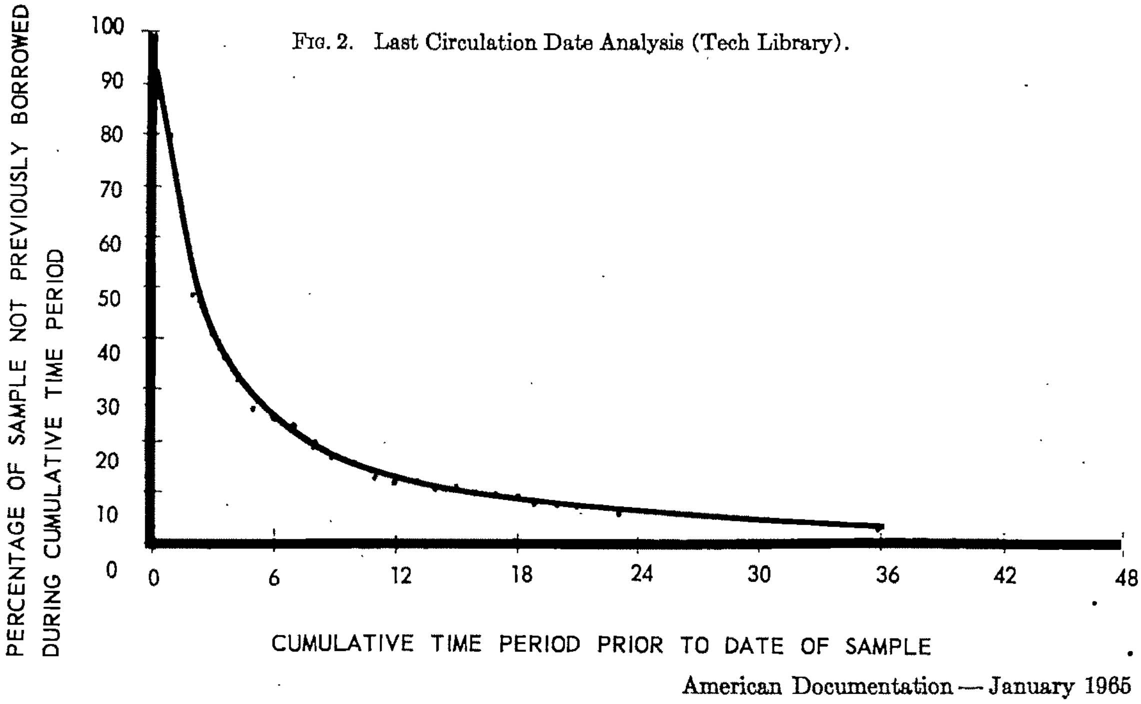


FIG. 2. Last Circulation Date Analysis (Tech Library).

If this calculation is continued, that is, for each month after the eight year point in time, we can then compute or estimate the size of the core collection.

This calculation was made for the Deering Library and for the Technological Institute Library at Northwestern, and it was found that approximately 25% of the current holdings of the Tech. Library should satisfy over 99% of the current circulation requirements. Similarly, a figure of about 40% was obtained for the Deering Library.

Thus, in summary, the basic approach centers around the use of the cumulative distribution of the previous circulation date. Assuming that this distribution represents typical circulation for the given library, the 99%-tile or 99.5%-tile position is then determined and this point in time is considered to be the cutting point for thinning the stacks. Such a cutting point should, if the system works properly, provide a collection that will satisfy over 99% of the current circulation requirements. If the resulting distribution and previous monthly circulation figures are then multiplied together for each of the corresponding months since the point in time established by the 99.5%-tile point, one can then calculate the expected size of the core collection. According to the very limited data and small samples taken at the Technological Institute Library and the Deering Library, it was predicted that such core collections satisfying over 99% of the circulation requirements at each of these libraries were made up of books that represented approximately 25% and 40% of the present holdings. Thus, it is inferred that 60 to 70% of the holdings might be removed to temporary storage, to an inter-library center, or to another limited access area.

Certain very questionable assumptions have been made in the above approach. The major assumption is that the cumulative distribution function for the current circulation was descriptive of circulation over time during the life of the library or at least back to the 99.5%-tile position. Such an assumption seems reasonable but cannot be accepted as valid without information taken over time for the libraries concerned. However, if the cumulative distribution correctly represents the circulation requirements of the users and if this distribution is typical over time, then the approach is reasonable.

It should be noted that no attempt has been made to evaluate the effect of this approach on in-stack use or browsing. Fussler (1961) concluded that generally those types of books having high circulation use also had high in-stack use. These relationships should be explored further. There is also the question of the value of the given volume to the user. The needs of a research scientist working on an urgent problem of national importance certainly exceed the needs of a freshman doing research on a homework problem. This study has, by its very design, assumed equal values in both of these situations. It would be possible, however, to categorize the users and develop user-differentiated, cumulative distribution functions and to then apply these to operating policies.

• Evaluation of Method

In an attempt to evaluate the technique in actual practice, an analysis was made of a one-week circulation sample in the 820's and 830's of the Deering Library. In both of these areas, the less-than-1% point occurred at the 13 year mark. Summing the products of monthly circulation and the percentage not previously circulated since the 13 year period began, it was possible to calculate or predict a core collection that should satisfy over 99% of current circulation requirements. In the case of the 820's, it was predicted that a core collection of 26,428 volumes existed in a 32,000 volume collection of the 820's. Thus, in effect, if this decision rule were applied, the new core collection is predicted to be 82.6% of the present holdings. Similar computations for the 830's showed that the core collection was approximately 54.2% of the present stack holdings. Thus, core collections of 82.6% and 54.2% were predicted for the 820's and 830's through the use of the decision rule and procedure described above.

As a test of the validity of this approach, a 1.7% random sample of those books in the stacks in the 820's and 830's was made and the last circulation date recorded. This last circulation date was then examined and the decision rule applied that books should be removed from the collection if the last circulation date occurred prior to the past thirteen years. Application of this decision rule to the sample taken from the books in the stacks provided a sample core collection of 79.6% and 54% for the 820's and the 830's. These values can now be compared to the predicted values of 82.6% and 54.2%. It is felt that the extreme closeness of these two sets of figures arrived at by two different methods is really a result of chance, but it does indicate a very strong possibility that the technique is a valid one.

Further research is planned by the author to evaluate this technique as a method for thinning stacks and for determining the requirements for multiple copies or non-circulating copies, etc. The latter types of categories could be predicted or determined by considering the cumulative frequency distribution of last circulation data for current circulation and pre-defining or pre-deciding the given percentage of current circulation requirements to be satisfied by a non-circulating or multi-copy core collection. Thus, it might be decided as a policy statement that a non-circulating core collection would be such that it would satisfy 90% of the circulation requirements. Examination of figure 2 would then indicate that the decision rule for selecting those volumes that are within this 90% satisfaction category would be those books that had circulated one or more times in the past 18 months. By calculations similar to the above, the size of this collection could be predicted.

Another representation of the above is a plot of percentage of user requirements versus percentage of holdings used to satisfy the given user requirements. This

plot is based on a sampling of previous circulation date for books circulated and for books in the stacks. It is possible to obtain from this plot information of the form: "X% of the circulation requirements are satisfied by Y% of the library's holdings."

Data Collection

The following are details of the procedures used to collect data. The book cards for each day's circulation were saved, and the last circulation date for each was recorded with subsequent punching in tabulating cards. The first three digits of the class number and the first three digits of the accession number were also recorded and subsequently punched. The latter information permitted analyses within subject areas and also an analysis by "age" of the book (that is, the age or the time since the book entered the library and received an accession number.) These cards were then sorted by last circulation date and listed on an accounting machine. Using this printout, it was possible to locate the 99%-tile as well as grouping by time periods of monthly and yearly intervals. This information was then plotted as shown in figures 1 and 2. In the few cases where the current circulation date was the only entry on the book card, some decisions were necessary because this indicated either a new book card or that this was the first time the book had circulated. Most of these cases could be resolved by comparing the age of the book from the accession number with the apparent age of the card from its physical appearance.

The last circulation date for books in the stacks (820's and 830's) was determined by physical examination of the book card in the jacket of the book. The sampling procedure used was to take the second book from each shelf in the stacks for the subject area sampled. This provided a 1.7% sample of the books. There is some question as to whether or not this sample is truly random, considering the fact that the books are placed within subject categories when the Dewey Decimal classification system is used. Further research on sampling procedures is needed. The information taken from the book cards of books in the stacks was then examined to determine the number of volumes that have their last circulation date on or before the 99%-tile date for the circulated books. Knowing the number of volumes in the sample, it was then possible to determine the percentage of the holdings having a last circulation date older than the one specified. This percentage indicates the proportion of the stack holdings that would be removed if the 99%-tile date (predicted from circulation requirements) were used to thin the holdings.

Conclusion

The strategy outlined above described a tool that may be of value for thinning the stacks according to the criteria of user needs. The resulting stack collection would by definition satisfy well over 99% of the user circulation requirements and would be of minimum size to satisfy these requirements. The decision rules to apply are relatively simple and could be applied by low-cost help rather than the use of a professional librarian for stack thinning. The thinning strategy by its very nature in reflecting user requirements would keep in the holdings those books that are used frequently and needed by the users. It would not select out books that might be very old and yet still used frequently. Similarly, it would select out those books that are relatively new and yet very infrequently used or of little value to the library's user population. Using this method, it is also possible to determine those volumes that should have multiple copies. The effective result of this approach is that there may be a quantitative way of maintaining the holdings of certain libraries at a reasonable level and a way of reducing the disappointment of the user who is looking for a specific volume.

Bibliography

FUSSLER, HERMAN H., and SIMON, J. L. *Patterns in the Use of Books in Large Research Libraries.* Chicago: University of Chicago Library, 1961.

GNOZA, E. "Survey of the Circulation Department, Charles Deering Library." Unpublished paper. Evanston: Deering Library, 1951.

KILGOUR, F. G. "Recorded Use of Books in the Yale Medical Library." *American Documentation,* vol. 12 October, 1961, pp. 266–269.

KILGOUR, F. G. "Use of Medical and Biological Journals in the Yale Medical Library." *Bulletin of the Medical Library Association,* vol. 50, No. 3, July 1962.

KILGOUR, F. G., and FLEMING, THOMAS P. "Moderately and Heavily Used Biomedical Journals." *Bulletin of the Medical Library Association,* vol. 52, No. 1, January 1964.

MCGAW, H. F. *"Policies and Practices in Discarding." Library Trends,* January 1956, pp. 269–282.

MEIER, R. L. "Communications Overload: Proposals from the Study of a University Library. *Administrative Science Quarterly,* March, 1963, pp. 521–544.

ROY, ROBERT, et al. "Progress Report on an Operations Research and Systems Engineering Study of a University Library, (NSF G–N–31)." Unpublished report. Johns Hopkins University, Baltimore, 1963.

TRUESWELL, RICHARD W. "User Behavioral Patterns and Requirements and Their Effect on the Possible Applications of Data Processing and Computer Techniques in a University Library." Unpublished doctoral dissertation. Northwestern University, Evanston, 1964.

TRUESWELL, RICHARD W. "Two Characteristics of Circulation —and Their Effect on the Implementation of Mechanized Circulation Control Systems." To be published in *College and Research Libraries,* vol. 25, No. 4, July, 1964.

TRUESWELL, RICHARD W. "Determining the Optimal Number of Volumes for a Library's Holdings." Presented at the joint meeting of Operations Research Society of American and the Institute of Management Sciences, Minneapolis, Minn., October 7-9, 1964.

Matching of Question and Answer Terminology in an Education Research File[1]

The hypothesis that relevant answers in a computer output may be distinguished from peripherally relevant and nonrelevant answers by the relative frequency with which question words and other related terms match words in the title, abstract, and machine index to these answers is investigated, using a sample of 14 questions searched against a pilot information file for education research. The frequencies of (1) significant words of the question, (2) words related thesaurally to the question words by the semantic code, and (3) words added to the original statement by the question analyst as answer index terms are used to determine empirically a function indicating the indexing and evaluating effectiveness of each. Similarly, the frequencies of question words in titles, conventional abstracts, and machine indexes are used to compare the indexing and evaluating effectiveness of these three forms of index.

JEAN TAGUE

Center for Documentation and Communication Research
School for Library Science
Western Reserve University
Cleveland, Ohio

Successful information retrieval depends on a matching of question and answer terminology. It is well known that the words a questioner uses to state a request are not necessarily the words which will have been assigned as indexing terms to answers acceptable to the questioner. As part of a study of the effectiveness of the pilot information service for educational research at the WRU Center for Documentation and Communication Research, an attempt has been made to classify the relationships between question words and words used to index relevant, peripherally relevant, and nonrelevant answers.

In most studies of retrieval systems the term "relevance" is considered to be an undefined concept or is given some such vague definition as "a relevant document is one which answers the question" or "a relevant document is one which satisfies the questioner." In these cases, relevancy is being determined subjectively by the questioner or other evaluator.

It is also possible to define relevance objectively; i.e., a relevant document is a member of a set defined by certain objective and possibly quantifiable characteristics. For example, in Cleverdon's study (1) the relevant document is the source document for the question; in other words, trial questions submitted to the system are phrased from documents known to be in the system. Another type of objective — and, in this case, quantifiable — relevance is the "relevance number" of Maron and Kuhns (2). Objective criteria of relevance are usually based on a special situation, such as questions with predetermined answers or an intimate and detailed knowledge of user needs, as in a research library serving a small clientele.

The education research file, established during 1961 and 1962 under contract with the U. S. Office of Education, is designed to serve the informational needs of educational research workers, administrators, and practitioners throughout the country. At least initially, because the file will be used by a diverse group whose interests range from experimental psychology to practical classroom techniques, few assumptions could be made about the characteristics of documents questioners would consider relevant to their request. It therefore appeared useful to obtain subjective assignments of relevance from the questioner and then to determine the objective properties of the documents selected as relevant.

Specifically, the following questions were asked:

1. Are answers considered relevant by the questioner more likely to be indexed by terms used in stating the question than answers considered peripherally relevant or nonrelevant?
2. Are relevant answers more likely to be indexed by words related thesaurally to the question words by the semantic code (the system of terminological con-

[1] Research supported by the Office of Education, U.S. Department of Health, Education and Welfare, Cooperative Research Project No. 1743.

trol used in the first stage of the project) than peripheral or nonrelevant answers?

Are relevant answers more likely to contain words which the question analyst has added to the original questioner's statement than peripheral and nonrelevant answers?

4. What are the differences in the distributions of these three types of words and of all words included in the search program (i.e., the form of the request used in searching the file) in relevant, peripheral, and nonrelevant answers?
5. What are the differences in the distribution of question words in titles, abstracts, and machine indexes ("telegraphic abstracts") for relevant, peripheral, and nonrelevant answers?

Our purpose in attempting to isolate some property or properties of relevant, as opposed to peripheral and nonrelevant, answers is to simplify and expedite the identification of these answers within the present retrieval system. If relevancy is based on a property external to the indexing and retrieval system, such as total number of answers scanned or order in which answers are scanned, general rules for identification of these answers will be difficult to formulate. As a first hypothesis, therefore, it was postulated that some linguistic property — some characteristic of the language of an answer — distinguishes relevant answers. In more specific terms, this hypothesis may be stated as follows: relevant answers are distinguished from peripheral and nonrelevant answers by the relative frequency with which question words and words with varying degrees of relationship to the question match words in the title, abstract, and machine index to these answers.

Three degrees of relationship will be considered:

1. The question words themselves;
2. Words related thesaurally to the question words by the information retrieval system;
3. Words related to the question words by the question analyst, independent of the system, through his knowledge of the file or subject area.

Sample for the Study

A total of 1,214 abstracts, representing answers to 14 questions,[2] were evaluated by the original questioners and returned to the Center. These questions had been selected from the more than 400 gathered during the initial phase of the project from educators, research workers, and other interested individuals in response to a widely circulated announcement of the pilot information service. They were searched against the pilot file of 4,837 documents. A brief statement of each question and the total number of answers sent to the questioner follows:

[2] The original sample contained 24 questions, but only 14 questioners evaluated and returned answers.

No.	Question statement	No. of answers
1.	High school dropout as related to juvenile delinquency	6
2.	Pupil-directed learning (independent study or discovery)	61
3.	Influence of parents on their children's school performance	246
4.	Decision-making	174
5.	Education of migratory children	24
6.	Role of teaching machines in treating mental illness	8
7.	Student interests and how they relate to achievement	21
8.	Use of audiovisual aids in teaching deaf children	3
9.	Meaning of achievement test results in terms of actual performance	36
10.	Programmed learning as compared with other teaching methods	334
11.	Teaching of film (i.e., film appreciation, literacy, comprehension)	53
12.	Teaching of concepts by inductive or deductive techniques	202
13.	Effectiveness of learner participation in teaching motor skills by film	39
14.	Relationship of scholastic achievement to success on the job	7

Evaluation categories were defined as follows:

Relevant: an article which has direct bearing on the question
Peripheral: an article of limited or possible use in relation to the question
Nonrelevant: an article which has no bearing on the question

Only one relevance assessment was made for each answer — that by the original questioner. It would, of course, be of interest to determine the variations in relevancy assessment among a group of evaluators, and a study of this aspect is now underway with the education research file. However, what is investigated here is the questioner's verbal statement of his informational need. To obtain two equivalent evaluations of a set of answers, two questioners who had independently submitted identical verbal statements would be needed; this situation did not, in fact, occur among the sample questions submitted.

Another variable not considered here but currently under investigation with the education research file is the order in which answers are examined. It is possible that if answers are examined in two different orders, two different assessments of relevance will be obtained. However, here again, to test this hypothesis, more than one evaluator is needed for each set of answers. When it is required that evaluators be the original questioners, such a test is not feasible.

• Operation of the Pilot System

Much has already been written about the semantic code and other distinctive features of the retrieval system developed at Western Reserve University. The unit operations as applied to the education research file are described in detail in the initial report on this project (3). However, for convenience, the main features will be briefly reviewed.

The educational research file, at the beginning of this study, consisted of 4,837 documents published in the years 1952 to 1962, "documents" being periodical articles, institutional or governmental research reports, dissertations, books, and chapters of books.

Two types of abstracts are written for each document. The conventional abstract is an English language summary of the article, giving bibliographic data and the significant contents of the article. It is designed to serve as the "product" of a search and a guide to the questioner in deciding whether or not he wishes to consult the original article. The telegraphic abstract analyzes the subject content of a document in a form amenable to machine searching. It consists of a list of the significant words of the article organized into syntactical relationships by "role indicators" and "punctuation." The words in the telegraphic abstract are keypunched, matched against an English word–semantic code dictionary file, and encoded automatically, unless the word is new, in which case a code is assigned.

The semantic code serves:

1. To make explicit the thesaural relationships of the relatively uncontrolled terminology of the telegraphic abstract;
2. To convert the natural language of the telegraphic abstract into machine-readable form.

A semantic code is composed of:

1. Four-letter semantic factors which express the generic concepts included in the English term and thereby relate it to other terms representing some or all of the same concepts;
2. Four-digit numbers distinguishing words containing identical semantic factors from one another.

Some examples may clarify this description:

English word	Code	Concepts represented by semantic factors
Teacher	PAPL SUCH 2002	PAPL — person
Professor	PAPL SUCH 2104	SUCH — educating, teaching
Curriculum	PALC. 1003 SYCH 2138	PALC — policy, plan PALC. 1003 — program SYCH — taught, learned
Educational TV	MACH SUGT SUND SYCH 4100	MACH — device, machine SUGT — making use of sight SUND — making use of sound

The encoded punched cards are used as input for the magnetic tape file of telegraphic abstracts which is searched by the GE 225 computer for answers to informational requests.

• Procedure for Asking Questions

In order to ask a question of the tape, a program is written which requires various configurations of semantic codes, role indicators, and punctuation. Permissible connectives includes conjunction (and), disjunction (or), and complementation (but not), so that the complete program may be represented by a Boolean polynomial in which the terms are semantic codes and/or role indicators. The coded words in the program include one or more of the following types:

1. Significant words of the question;
2. Words related thesaurally to the question words by the semantic code — i.e., words including at least one of the same semantic factors as the related question word;
3. Words not related to question words by the semantic code, but known by the question analyst, through intuition or knowledge of the file, to be related to the question words.

We shall call these three types respectively question words, semantic code references, and program additions.

To clarify this categorization, the words of the first question are classified below according to type:

	Word	Code	Symbol
Question words:	High school	LACN SUCH 2101	A_1
	Dropout	T–RM.1131	B_1
	Juvenile	PYPL.1317	C_1
	Delinquency	DVDD PASS 2131	D_1
Semantic code references:	Junior high school	LACN SUCH 2102	A_2
	Secondary school	LACN SUCH 2103	A_3
	Dropping out	T–RM.1131	B_1
	Termination, terminating	T–RM.1001	B_2
	Discontinuance	T–RM.1010	B_3
	Crime	DADD PWSS 2393	D_2
Program additions:	Withdrawal, withdrawn	R–ML.1002	B_4
	Continuance	CANT.1010	B_5

The program may be written in Boolean polynomial form using the symbols at the right:

$$\{ (A_1+A_2+A_3)\cdot(B_1+B_2+B_3+B_4+B_5)\cdot C_1\cdot(D_1+D_2) \}$$

• Analysis of Evaluated Answers

The frequency of each type of word — question words, semantic code references, and program additions — in relevant, peripheral, and nonrelevant answers was computed and used to determine the effectiveness of each in:

1. Locating relevant answers (i.e., as indexing terms);
2. Indicating the relevance of answers (i.e., as evaluating terms).

TABLE 1. Comparison of Question Words, Semantic Code Reference, and Program Additions Indexing Terms

No. of Words	Probability of occurrence (Recall factor)											
	All words			Question words			Semantic code refs.			Program additions		
	Ref.	Per.	Nonrel.	Rel.	Per.	Nonrel.	Rel.	Per.	Nonrel.	Rel.	Per.	Nonrel.
1 or more	1.00	1.00	1.00	.86	.71	.66	.25	.30	.47	.71	.45	.39
2 or more	.87	.60	.62	.46	.43	.35	.04	.05	.04	.39	.12	.12
3 or more	.60	.32	.26	.14	.09	.04	0	.01	0	.08	.03	.05
4 or more	.31	.15	.11	.02	.02	0	0	0	0	.01	.01	.02
Mean	2.92	2.16	2.02	1.48	1.25	1.05	.29	.36	.51	1.19	.61	.58

The effectiveness of a word A as an indexing term is rated by the ratio of number of relevant answers indexed by A to total number of relevant answers for that question. This ratio is often called the "recall" factor. One may extend the concept of indexing effectiveness from single words to types of words — the types of words, in this case, being question words — semantic code references, and program additions. With word types, however, one must consider not just a single ratio, but a frequency function plotting number of type A words against the observed frequency of this number of type A words in relevant answers; i.e.,

$$\text{Probability of } \chi \text{ type } A \text{ words in relevant answer} = \frac{\text{No. of relevant answers with } \chi \text{ type } A \text{ words}}{\text{Total no. of relevant answers}}$$

The distribution of words of type B may similarily be plotted and compared with words of type A. If the mean number of words of type B is greater than the mean for type A, and if this difference is significant, then the former are said to be more effective as indexing terms.

The "relevancy" factor — i.e., the ratio of the number of relevant answers containing a term A to the total number of answers containing term A — indicates the evaluating effectiveness of this term, since it answers the following question: if a questioner receives all answers containing word A, what proportion of these will be relevant? This concept may also be extended to types of words by a function plotting the number of words of type A against their relevancy factor. The function will not be a frequency distribution, since the probabilities do not total one. However, if the curve for words of type A is everywhere above that for type B, we may conclude the former are more effective as evaluating terms.

In a similar manner, the relative effectiveness of question words in titles, conventional abstracts, and telegraphic abstracts as indexing and evaluating terms may be compared on the basis of frequencies in relevant answers.

• Results

QUESTION WORDS, SEMANTIC CODE REFERENCES, AND PROGRAM ADDITIONS

The columns for relevant answers in Table 1 show the relative effectiveness of question words, semantic code references, and program additions as indexing terms for our sample of 1,214 answers, made up of 474 relevant, 285 peripheral, and 455 nonrelevant answers. Table 2 shows the relative effectiveness of question words, semantic code references, and program additions as evaluating terms. The mean number of matches for each type indicates that question words are most effective as indexing terms, followed by program additions. Semantic code references appear to have comparatively little indexing value. Table 2 presents the surprising result that up to combinations of two or more, the program additions are more effective than question words in evaluating answers.

Table 1 indicates that matching of words in program and telegraphic abstract, regardless of type, is greater for relevant answers than for peripheral or nonrelevant answers. Peripheral answers are similarly distinguished from nonrelevant answers, but the differences are not so great. The distribution of each type of word and of total words in relevant, peripheral, and nonrelevant answers can be compared by applying a Chi square analysis to each pair of distributions. Such an analysis determines whether relevant answers are characterized by a particular type of language. At the .01 level, the frequency of total words in program, question words, and program additions was found in each case to be greater in relevant answers than in peripheral or nonrelevant answers; the frequency of total words, question words, and program additions did not differ significantly between peripheral and nonrelevant answers. The frequency of semantic code references did not differ between relevant and peripheral answers but is significantly greater in nonrelevant than in relevant or peripheral answers.

TABLE 2. Comparison of Question Words, Semantic Code References and Program Additions as Evaluating Terms

No. of Words	Probability of Relevance (Relevance Factor)			
	All Words	Question Words	Semantic Code Refs.	Program Additions
1 or more	.39	.45	.29	.52
2 or more	.47	.44	.39	.66
3 or more	.57	.61	...	.58
4 or more	.57	.78	...	.42

t relevant, peripheral, and nonrelevant rimarily in the total number of words :n question program and telegraphic ab- an in a particular type of word match. differences more exactly, let us define a which will measure the distance between heral and peripheral and nonrelevant answ... total word matches, question word matches, etc. The distance between relevant and nonrelevant answers in terms of question word matches (d_{QRN}) is defined as

$$d_{QRN} = \sqrt{\sum_{x=0}^{4} [P(XQ/R) - P(XQ/N)]^2}$$

where:

P(XQ/R) is the probability of X question words matching in a relevant answer

P(XQ/N) is the probability of X question words matching in a relevant answer

This formula was suggested by that used by Osgood (4) in measuring the similarity in meaning of two concepts. Distances can vary from 0 to $\sqrt{2}$. The distance will be 0 if the probability of words matching between question and answer is exactly the same in relevant and nonrelevant answers for all degrees of matching. The distance will be $\sqrt{2}$ if, for example, the probability that 1 word matches in relevant answers is 1 and the probability that 0 words match in nonrelevant answers is 1.

One may similarly define the distances between relevant and nonrelevant answers in terms of semantic code references, program additions, and total words; and the same distance between relevant and peripheral answers. The results are as follows:

Distance between relevant and nonrelevant answers in terms of:

Question words	.25
Semantic code references	.22
Program additions	.42
Total words	.52

Distance between relevant and peripheral answers in terms of:

Question words	.16
Semantic code references	.05
Program additions	.38
Total words	.42

Taking Tables 1 and 2 in conjunction, one can summarize the results which would have been obtained for this sample with various types of question programs, as shown in Table 3. Some general conclusions may be drawn from the data of this table:

1. Question words are most useful in locating relevant answers, but cannot be employed exclusively if completeness is desired, since 14% of the relevant answers contained no question word at all.

TABLE 3. Results of Various Types of Program on Search Output

No. of words required by program	Type of words required by program	% of Relevant answers retrieved	% of Output which will be relevant
1 or more	Any type	100	39
	QW	86	45
	SCR	25	29
	PA	71	52
2 or more	Any type	87	47
	QW	46	44
	SCR	04	39
	PA	39	66
3 or more	Any type	60	57
	QW	14	61
	SCR	0	..
	PA	08	58
4 or more	Any type	31	57
	QW	02	78
	SCR	0	..
	PA	01	42

2. Semantic code references to question words are least effective in programs either in retrieving relevant answers or in distinguishing relevant from peripheral and nonrelevant answers.
3. The greatest drop in percent of relevant answers retrieved, considering words of all types, is between combinations of two and three words. If at least two words are required, 87% of the relevant answers will be retrieved; if at least three words are required, 60% of the relevant answers will be retrieved.
4. The greatest drop in percent of relevant answers retrieved by question words occurs between combinations of one and two words. If at least one word is required, 86% of relevant answers will be retrieved; if two or more words are required, 46% of relevant answers will be retrieved.
5. The program additions appear to be the most effective evaluating terms, at least in combinations of one and two.

TITLES, CONVENTIONAL ABSTRACTS AND TELEGRAPHIC ABSTRACTS

In comparing words in titles, conventional abstracts, and telegraphic abstracts, it was considered advisable to admit any grammatical variant of a question word (e.g., teaching, teacher, taught, teachable), since, in consulting a conventional index it is likely that a questioner would "look up" under all forms. The relative effectiveness of titles, conventional abstracts, and telegraphic abstracts as indexes is shown in Table 4; as evaluating tools, in Table 5. From the means of the relevant answer columns in Table 4 it is apparent that words in the conventional abstract are the best indexes to relevant answers. However, differences between the two types of abstracts are small compared with the difference between these and titles, the title words having a much lower indexing effectiveness. On the other hand, title words are most effective as

TABLE 4. Comparison of Title Words, Conventional Abstract Words, and Telegraphic Abstract Words as Indexing Terms

Number of Question Words	Probability of occurrence								
	In title			In conventional abstract			In telegraphic abstract		
	Rel.	Per.	Nonrel.	Rel.	Per.	Nonrel.	Rel.	Per.	Nonrel.
1 or more	.65	.39	.37	.88	.63	.61	.90	.80	.81
2 or more	.22	.11	.07	.60	.27	.30	.54	.49	.45
3 or more	.05	.02	0	.25	.07	.08	.19	.15	.13
4 or more	0	0	0	.04	0	0	.02	.03	.02
Mean	.92	.54	.45	1.77	.97	.91	1.65	1.47	1.41

evaluating terms — an expected result since these generally are the most significant of the indexing terms. The lower evaluating effectiveness of the telegraphic abstract is due partly to the fact that the questioners do not see the telegraphic abstracts, only titles and conventional abstracts. This same situation may explain the high frequency of question words in the telegraphic abstracts of peripheral and nonrelevant answers. If an appropriate term appears in the conventional abstract, the questioner will see it and call the answer relevant; if the same term appears in the telegraphic abstract only, he may well call the same answer nonrelevant.

A Chi square test of the difference in distribution of words in relevant *vs.* peripheral and nonrelevant and peripheral *vs.* nonrelevant answers at the .01 level shows that for all three indexes — titles, telegraphic abstracts, and conventional abstracts — the frequency of question words is greater in relevant than in peripheral or nonrelevant answers but does not differ significantly between peripheral and nonrelevant answers. Thus it may be concluded that the frequency of question words in titles, conventional abstracts, and telegraphic abstracts distinguishes relevant from nonrelevant answers. The distance function defined earlier may again be used to indicate the relative strength of these differences:

Distance between relevant and nonrelevant answers in terms of:

Titles	.32
Conventional abstracts	.44
Telegraphic abstracts	.14

Distance between relevant and peripheral answers in terms of:

Titles	.28
Conventional abstracts	.45
Telegraphic abstracts	.12

Differences in the telegraphic abstract are slight; the greatest distinctions are in the conventional abstracts.

For this sample, the following conclusions may be drawn:

1. Titles alone have a low indexing effectiveness. If titles are searched for one or more question words, at most 65% of the relevant answers will be retrieved, whereas if telegraphic abstracts are searched, 90% of relevant answers will be retrieved.
2. Indexing terms which appear in the telegraphic abstract but not in the conventional abstract are of little value.

• Conclusions

This particular sample might be criticized on two grounds:

1. It is too small;
2. Each answer is given equal weight in the analysis, but the weight assigned to each question varies in proportion to the number of answers which it retrieves (ranging from 3 to 334).

Unfortunately, the delays inherent in sending answers outside the Center for evaluation made it inconvenient to increase the sample size when it became apparent that all of the original 24 questioners were not returning answers. Because of these disadvantages, it is felt that this study must be regarded as a preliminary investigation, whose main interest lies more in the method than in the actual results. For more conclusive results, the procedures should be extended to a larger sample.

However, the following conclusions might be drawn, even from this imperfect sample:

In programming questions, the number of words required in combination is of greater significance than the type of words required as far as retrieval of relevant as opposed to peripheral or nonrelevant answers is concerned. In other words, it is the number of words rather than the relationship of the words to the original question which characterizes relevant answers.

The related words which the question analyst adds to the original statement of the question are useful as in-

TABLE 5. Comparison of Title Words, Conventional Abstract Words and Telegraphic Abstract Words as Evaluating Terms

No. of Question Words	Probability of Relevance		
	In Title	In CA	In TA
1 or more	.52	.48	.42
2 or more	.61	.57	.44
3 or more	.65	.65	.49
4 or more	...	...	...

dexing and evaluating terms, although their indexing effectiveness is lower than that of question words.

The semantic code leads from the question words to very few additional indexing terms, and those which it selects are more likely to retrieve peripheral and non-relevant than relevant answers. Theoretically, the semantic code should have related to question words not only the words designated as semantic code references but also those designated as program additions. That it did not might be attributed: (1) to the fact that the semantic code for words included in the questions did not provide a sufficient number of thesaural associations; or (2) to the limitations of any system in which thesaural associations are preassigned. From this latter point of view, relevant thesaural associations cannot be specified for words in isolation, only for words as they appear in the context of a question. Which of these two explanations is the true one can be discovered only through further investigation.

Because of their indexing effectiveness, titles need not be recorded on tape as separate indexing entities.

All of the information in the machine index should be made available to the questioner in the search "product" which is sent to him.

References

1. CLEVERDON, C. W. 1962. *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems.* College of Aeronautics, Cranfield, England.
2. MARON, M. E., and KUHNS, J. L. 1960. "On Relevance, Probabilistic Indexing and Information Retrieval." *Journal of the Association for Computing Machinery,* v. 7, pp. 216–244.
3. REEVES, PAMELA W., *et al.* 1962. *The Library of Tomorrow—Today.* Western Reserve University, Center for Documentation and Communication Research, Cleveland, Ohio.
4. OSGOOD, CHARLES E., SUCI, GEORGE J., and TANNENBAUM, PERCY H. 1957. *The Measurement of Meaning.* University of Illinois Press, Urbana, pp. 90–91.

BRIEF COMMUNICATIONS

A Boyle's Law for Indexer Consistency

Everybody knows that with large document collections only the naive or masochistic searcher attempts at all times to retrieve all documents pertinent to his interests. The whole point of the so-called "negotiation of the question" is designed to reduce the "all" to some manageable subset. Many systems, both mechanized and manual, employ post-editing of retrieval output, as well, to further reduce the mass of material generated by broad requests.

This well-known fact should have had an influence on the design of experiments on indexer consistency. To my knowledge it has not; and the point of this note is to indicate an experimental design which might better account for the fact that searchers are frequently satisfied in spite of the well-known imperfection of indexers.

What I want to suggest is that one need not expect that two indexers (or one indexer at different times) will, for any *given* document, assign either the same or very similar sets of terms. What one should expect is that for any homogenous *collection* of documents any two indexers (or one, etc.) will assign terms with roughly the same relative frequencies (or because of differences in depth, a high correlation between rank orders of relative frequencies) and that a document set retrieved on the basis of one indexing will be equally satisfactory to that set retrieved by the other.

It is a mistaken notion, I think, to expect that for any given request there is always a "best" answer in terms of a unique set of documents. Most collections, especially large ones, contain considerable redundancy of document content. Any answer which selects some representative (even if sometimes large) sample from a homogenous redundant collection is probably as good as any other. Hence, the phrasing above of "equally satisfactory."

The more demanding among us might expect that the two selected document sets would almost entirely overlap but this probably is just hopeful. It would, however, prove both interesting and important if there were significantly more consistency between document sets retrieved than between index terms assigned.

The view of the problem is not only more generous to indexers but also helps to account entirely for the otherwise surprising results of Resnick[1] and partly for the results of Lancaster and Mills[2] with respect to their "Relevance 2."

Similar views in another context have been suggested elsewhere.[3]

TERRY R. SAVAGE
Datatrol Corporation

References

1. Resnick, A. "Educational Requirements for Indexers in a Selective Dissemination System," *Automation and Scientific Communication,* ADI Annual Meeting, 1963, p. 163.
2. Lancaster, F. W., and Mills, J. "Testing Indexes and Index Language Devices: ASLIB Cranfield Project," *American Documentation,* v. 15. no. 1 (Jan. 1964), p. 4. See esp. p. 8.
3. Rath, G. J., Resnick, A., and Savage, T. R. "The Formation of Abstracts by the Selection of Sentences," *American Documentation,* v. 12, no. 2 (Apr. 1961), p. 139. See esp. p. 143.

Factors of Selection in Defining Science Information for the Federal Government

A definition of science information is not essential to effective information handling, but any evaluation of information handling implies the use of a definition.

In order to make science information activities "visible, audible, and accountable," and to evaluate the interrelations of information activities, there is a need to identify and characterize scientific information projects with details of budget, staff, mission, procedures, and achievements. As a preliminary to such an inventory and evaluation, it is necessary to define what is to be selected for examination; moreover, if the definition is to be successfully applied it should be well understood and accepted.

In principle, a useful definition should exclude as much as possible and still embrace germane elements. In practice, few boundaries are satisfying.

The following comments aim simply to identify the areas of decision. These comments are intended to be evocative rather than dogmatic.

● Technology and Science

Although the range of a definition of scientific information would be relatively tight if it excluded information bearing on technology, the distinction is not easy to make. The community of interest is expressed in the title of the President's Office of Science and Technology. Nevertheless, OST requests that the distinction be attempted in analyzing science and technological information services.

● Training

If training is a component of a science information program, the subsidiary questions are: . . . does one include training in the sciences as well as in the employment of science information techniques? . . . should training in the sciences be counted only for science information specialists?

● Audience Characteristics

In defining science information, some prefer to exclude messages directed to the general public. Most agencies, nevertheless, regularly provide scientific information to the general public as well as to the practitioner and the scientist.

● Character of the Message

The character of the science information itself is a factor of selection. Slogans, emotional appeals, generalities, exhortations, imperatives, administrative regulations, legislation, essays of opinion, financial statements, and budget justifications, as well as reports of isolated events, appear to disqualify themselves. But there is no simple way to distinguish legislation, imperative writings, or administrative instructions from a recommended ordinance and code, a guide, a training manual, or a textbook.

● **Cut-off Points: Writing and Feedback**

The Department of Defense Glossary defines information as "the meaning assigned to data," something that emerges when data is organized. By this definition, one might exclude the process of seeking, collecting, and organizing the data. But at what point does data become meaningful: when the tabulation is completed? when the first draft of the report is outlined? when the message first appears? A pragmatic decision would be to begin with the writing of the message.

Definitions posed by the Office of Science and Technology go even further and exclude the writing of reports by scientists, except for reviews and other secondary literature. Otherwise, only the labors of professional writers and editors are counted, possibly on the grounds that there is no way of reckoning the time used by the scientist for writing a report.

Pragmatic considerations also would seem to rule cutting off the information process at the point where it reaches the user. Although the process is conceived as a cycle, with feedback from the user as an important moderating factor in transmission of information, there may not be much point to obtaining data on feedback, except as it provides evaluative information.

● **Ancillary Services**

To reduce the range of science information accounting, one might separate purely routine processes of information handling from the intellectual activities. "But is it practical to separate replication, packaging, mailing, mechanized searching and retrieval, data processing, coding, indexing, citing, typing or copying, binding, filing, proofing, or most forms of illustration from writing, editing, creative illustration, program design, system design, creation of a thesaurus, or annotation of a bibliography?"

It is equally difficult to separate travel expenses and living expenses associated with a science information project, such as a training grant or conference, from expenses which are only incidentally linked with such projects.

● **Character of the Media**

Attempts to categorize science information according to media exclude, as a rule, the so-called mass media. At the same time, closed circuit TV, specialized films, and scientific publications of all kinds may lie within the charmed circle. Specific science messages aimed at an open audience by TV and other general media might qualify.

● **Non-Contributory Activities**

The conference as a device for exchanging science information can scarcely be underrated. Still there remains the question: should some kinds of conference activities, such as housing costs, or certain kinds of conferences, meetings, workshops, and symposia be excluded?

Libraries are no more exempt than conferences from such scrutiny. One might ask to what extent they are used for storage or recreation rather than for information transfer.

● **Research in Information Transfer**

Research in information processes, including pertinent behavioral studies, clearly lies within the circle but organization of data about research other than research in information processes could be excluded on the principle that the information concerns costs, allocations, or similar administrative considerations, rather than science. It may be impractical to count administrative charges for science information except as a flat percentage of the agency total.

● **Association with a Document**

One concept would limit the scientific information inventory to activities directly associated with written reports and their distribution. Such a concept includes speeches or conversations directly based on a document or directly and immediately associated with a document in preparation, such as conference proceedings, and still excludes data gathering and informal talks. Then, what is a document—a letter to the editor, a chart, a film?

● **Accountability**

An activity difficult to account for in science information handling is the individual inquiry. It may be a practical necessity to exclude all individual answers except those performed routinely in specialized information centers.

● **Conclusion and Summary**

Anyone seeking to develop an inventory of science information activities is obliged to rule certain activities in or out according to the readiness with which they can be made accountable or characterized.

The main objective is to give a framework to an inventory which may be useful in evaluating, analyzing, and planning information programs, to make them visible and accountable. The definition ought to exclude, rather than include, as much as may be eliminated on principle without doing violence to the function of the inventory.

Marcus Rosenblum
Associate Special Assistant to the Surgeon General (Science Information)

Report on a Conference on the Education of Science Information Personnel Held July 27-28, 1964, in Cleveland, Ohio

Representatives from 16 university library schools and several federal agencies and industries met on July 27-28, in Cleveland, Ohio, to consider recent progress in the education of science information personnel. The invitational conference, sponsored by the Center for Documentation and Communication Research of the School of Library Science, Western Reserve University, was intended as a sequel to conferences on training convened at the Georgia Institute of Technology in October 1961 and April 1962.

No pronouncements, resolutions or grand plans issued from the Cleveland conference. Rather, the conference provided a medium for the generation of ideas and evoked confidence that progress has indeed been made in the formulation of university programs, the thinking behind their development, and attempts to solve implementation problems at various levels. A distinction was emphasized throughout between the science information specialist and the information scientist. This distinction not only influences the curriculum and staff at a university but also the expected role for which the graduate is trained. Interests of the group ranged from traditional librarianship through relatively standard clerical applications of machines to theoretical studies in mathematics and linguistics leading toward new methods of strong, transmitting, and communicating information. A consensus was *not* that each library school augment its activities to include information sciences research, but that the fruits of research be shared through the introduction of new points of view from the information sciences into the curriculum according to school mission and interests.

The librarian as mediator between man and graphic records is not expected to be replaced by machines in the foreseeable future. If his training is to be rooted in librarianship and leafed by contributions from the information sciences, by what mechanisms can this branching occur? Branching problems were discussed at four concurrent workshop sessions. At least three of these, on faculty, curriculum, and organization within existing academic structures, were different looks at a basic question: What does "information sciences" mean to me? Contributions from many disciplines and cooperation among many departments appear necessary. These include business administration or industrial management, computation centers, philosophy, mathematics, language, engineering, and sociology departments. Whether joint

faculty appointments or a self-contained entity is the more workable academic arrangement may have to be established empirically. Competent faculty members are hard to find, and sources from which they and students can be recruited are not clearly defined. Indeed, the fourth workshop, on recruitment of students, questioned the need to recruit. The question is particularly pertinent until objectives of training and curricula have been specified in relation to market needs and anticipated contributions from science and technology.

Much program enrichment through cooperative effort was recounted. Several inter-university efforts were described which involve exchanges of faculty (Columbia and Rutgers) and integrated programs to obviate course duplication (American, Catholic, and George Washington Universities). In Texas and the mid-West, working relationships with industry are shaping curricula and services to the community. Increasing use is being made of the guest lecture and the seminar series staffed by invitees from industry, government, and other university departments to supplement education programs. Proposals suggested at the Cleveland conference include the provision of arrangements for student on-the-job training and future working meetings among the schools addressed to particular operational problems. No attempts were made to standardize curricula or even to define essential elements of a curriculum. The cooperative efforts in being and planned seem, instead, to be directed toward a gradual reorientation in training, each school adjusting to its own university and market environments.

The conference served as a forum for self-appraisal as well as a medium for information exchange. Does the librarian have confidence in his profession? queried Dean Jesse Shera. What does an information scientist do? asked others. How does one guard against fads, gadgets, and vogue ideas attempting to enhance the traditional with the new? Can a workable solution result from viewing information sciences as librarianship in depth, as the contribution from research which is providing a theoretical framework for library science? If the information sciences are concerned with the exploitation and synthesis of all recorded knowledge, can the totality of education programs meaningfully span the library-to-information sciences spectrum and remain in communication with each other? What fundamental core of knowledge must be common to all? How can better understanding at the academic level be transmitted to the user and research levels?

Preliminary results of a Battelle Memorial Institute study of manpower requirements for scientific information activities were reported at the Cleveland conference. No specific sets of skills could be identified for particular operations; e.g., abstracting, indexing, performed by specific scientific information personnel. Responses to one set of questionnaires disclosed 8400 persons so employed, but 63% of the employers account for only one man-year apiece. Persons' bases for selecting science information work were more negative than positive. This may reflect the need for a common core of knowledge and suggest an explanation for shortcomings in information services and the lack of appreciation by management of the potential usefulness of science information activities.

Discussion of sponsorship of education programs led to comments on the philosophy of education practices *per se*. Most schools are considering master's degree programs. The Office of Education, HEW, can assist school librarian and Ph.D. candidate programs in which the major is library science. However, this does not reach the more abstruse areas of the information sciences. Some Ph.D. candidate research does get indirect support via federal agency sponsorship of information science research projects, but the basis for such expenditures is the project objective, not education or training. Unsponsored and unacknowledged academically is the pursuit of advanced training intermediate between the master's and Ph.D. level. An alarm was sounded at the conference against plunging headlong into Ph.D. programs and the substitution of a status symbol signifying accomplishment within the master-doctorate gap. The iconoclastic note was intended for all of academicia, to preserve the Ph.D. for research achievement. At a lower pitch, a query was raised into possibilities of starting and sponsoring training at the undergraduate level, to take advantage of the upgraded high-school education which is beginning to include such topics as set theory, relativity theory, and computer technology.

Dr. Stafford Warren, special assistant to the President for mental retardation, provided the look into the future at a dinner address on a national library of science system and network. His broad plan calls for establishing bases for compatibility among information systems and the training of an adequate number of persons along the library-information sciences spectrum for proper functioning of these systems. Ultimately, however, the success of any plan will depend on attractiveness of the profession to knowledgeable and imaginative individuals. On the programs at our universities the case rests.

Proceedings will be available from the conference organizer, A. Goldwyn of Western Reserve.

ROWENA SWANSON
Directorate of Information Sciences
Air Force Office of Scientific Research
Washington, D. C. 20333

Letters to the Editor

October 29, 1964

Dear Sir:

I have just received my copy of *American Documentation* and was particularly pleased with Isaac Welt's brief (too brief) editorial, "Information Science–Science Information." I do not suppose that everyone agrees always with [what] Isaac says at all times, but he often puts his finger right on the thing that is wrong. This time he has put in words what so many of us were saying at that late fiasco in Philadelphia. To quote Mr. Welt: "The theoretical and untried has, within our society and within the pages of this journal, gained ascendancy over the practical and the workable."

In Philadelphia, if one were not certain of the time, place and event, it would be possible to assume that one had stumbled into the annual convention of mathematicians. I do not want to knock theory, but it is about time that someone take a second look at what is being done to ADI.

In closing I would like to comment on Claire Schultz's suggestion in her letter that papers to AD be indexed by the authors. I am not certain that this will ever be adopted as it is just too practical, no mathematical formulae involved, and certainly too much in keeping with what documentalists, information scientists and librarians should be doing. In addition, I am not certain that some of the authors would be able to find a descriptor that would fit their paper.

Sincerely,
Charles W. Sargent, Ph.D.
Document Librarian

August 20, 1964

Dear Sir:

In response to the Editorial in *American Documentation,* July 1964—

"These men were seeking information, and, one suspects, were offered documents . . . and documents . . . and documents." What's so bad about that? Most of our customers are adults. In fact, they are educated adults. Furthermore, they are educated in the very topics dealt with by the documents. If anyone thinks he can tell an expert which is the best document or what the message is, he is either another and equal expert or an illiterate documentalist (which, unfortunately, is not a contradiction in terms).

The basic confusion still abides between information retrieval and document retrieval, in spite of Bar-Hillel's jeremiad in *American Documentation* of April 1957. The sooner we use two names for two things the better. The distinction between documents and information is not an accident, not a figment of the imagination of either a librarian or of a documentalist. It is a recorded fact of life, the way information comes packaged for retention. The distinction has existed since the first record and will by definition continue until the last one.

A document, a collection of unique marks, is not a stone, not an inadequate response to a certain type of request for assistance. It is not the wrong response as long as the librarian or the documentalist is capable of distinguishing between questions that can be suitably answered by documents and those that require substantive answers.

Perhaps we need another name for the first type of "question" to distinguish it from the second type, which accords more with the everyday meaning of "question." Fairthorne has suggested the lordly word "invoke" for the type of request which is properly responded to by the procurement of documents. Let us use it until a better (less lordly?) word is suggested or we are mathematically conditioned to refer to "question_1" or to "question_2."

Mrs. Lea M. Bohnert
C-E-I-R, Inc., and
The American University

September 18, 1964

Dear Sir:

I have just finished answering a six page questionary of the kind of indexing and information service available at my library. It was the ninth such inquiry received in the past three months; they average three every month. I suggest that the temporal and pecuniary resources expended in these efforts might be more usefully directed to a judicious analysis of the problems of documentation and the discovery of experimental methods for their practical solution.

Surveys of this type are generally inquiries into what equipment and what procedures are being used, and into what opinions may be held regarding more or less speculative matters. Regrettably, however, the meaning of the terms so frequently given in these questionaries is by no means clear. The poor fellow who must reply normally has many more questions than answers, viz. is there a difference between *document analysis* and *indexing? He* may recognize a distinction but does the man who sent the questionary to him recognize one? If through some coincidence they be speaking of the same thing, does the statistical analysis of the results in such a poll have any significance? This verbal quibbling illustrates the widespread lack of validation characteristic of these investigations.

Solutions to the knotty problems of modern documentation will not be found in solicitations for opinions or by generally canvassing for the state of the art. Answers to the complex difficulties of information processing are not often found by asking simple questions of distinguished authorities, and certainly not by inquiry of everyone on a mailing list obtained from who-knows-where.

Operative, effectual methods are secured by trying different methods of information input, storage, and retrieval, then analyzing the results obtained. The methods to undertake for trial are obtained by study of as many possible, logical routes as can be examined. There are very few principles, if any, that can be applied to the problems of modern documentation; it is an empirical technology and its answers must necessarily be chiefly empirical. It is to be regretted that this calls for a lot of labor, but the forward impulse of progress is scarcely achieved without hard work.

Faithfully yours,
Frank S. Wagner, Jr.
Technical Librarian

August 17, 1964

Dear Sir:

Without disrespect to your Mr. Brann or the National Library of Medicine's Mr. Taine, we feel constrained to note that the former's abstract of the latter's "The Medical Literature Analysis and Retrieval System" (*American Documentation,* April 1964) loses in translation to English some of the flavor of the paper as originally published. For the benefit of purists such as ourselves, permit us to submit an alternate abstract in the original language made by our consultant on librarianship, which has been circulating about our office since the paper first appeared:

"The Medical Literature Analysis and Retrieval System" by Seymour I. Taine, *Bulletin of the Medical Library Association,* v. 51–2, pp. 157–167, April 1963.

NLM's Taine gives those who should have intelligence about MEDLARS to understand that to obtain increased coverage of the current substantitive literature of the world up to totality, and proceeding from an embryonic systems concept aimed at reducing throughput time and exploiting a single inputting of information, the National Library of Medicine has achieved the in-house capability of operation of a machine configuration, itself capable of 150,000,000 characters including non-IM tags, through employing search specialists to output format parameters to provide a current awareness type service, and also employing typists to produce digitalized codes that are machine usable, all primarily to insure that the INDEX MEDICUS is outputted in a more interesting manner.

EDWARD WOOTEN
Managing Director
The Charles E. Wooten Company
Alexandria, Louisiana

Dear Sir:

Some years ago we had hanging in the Oregon State University Library Staff Room a series of prints from Pablo Picasso. These made good conversation pieces. One afternoon while at tea I remarked harmlessly, as I supposed, that the best thing about surrealist art is that it can mean to the viewer whatever he chooses to have it mean and that this will very likely be poles apart from what the artist had in mind, and from what other viewers will see.

On the opposite wall from where I sat hung a print depicting a series of markings, angular lines, and assorted colors, among which three squarish mechanical robot faces were somewhat discernible. This Picasso had chosen to call "Three Musicians." From the depths of my artistic illiteracy I asserted that these three so-called figures could, in these days of impending machine organization and control of knowledge, more aptly be dubbed "Three Librarians." This I thought would be appropriate since the robotic Librarians of the future, if they could not dispense service with a smile could easily do so to music. I also thought that Mr. Otis, the noted artist whom I had but just met a day or two before, could have handled this theme and created these three figures even more brilliantly than Picasso has done. Mr. Otis, for those who do not know him, is the fictional creation of the well known Portland, Oregon author Stewart Holbrook. Mr. Holbrook supplies learned artistic comment on a goodly number of the more famous Otis canvasses.

Little did I think that these passing remarks would send subconscious brain cells spinning. But they did for that night I had a dream. A strange looking mechanical man with squarish robotic face was rescuing an equally strange mechanical woman from a castle moat. I was relieved, but not for long, when the woman, in squarish mechanistic swim trunks, was safely on the castle bank. No sooner was she rescued, however, than the man, his square face sardonic and his eyes blazing, strangled her and flung her back into the moat. I was so startled by this unexpected denouement that I came wide awake thereby rescuing this fantasy from that strange limbo of the unconscious from whence it came and to which it would surely have hazily returned, little noted, or not at all, had I not awakened.

What, I mused, as I became aware of the similarity of these vanished and so recently vivid faces to Picasso's Musicians, what would Mr. Freud make of all this? Or Mr. Otis? Being as ignorant of psychiatry as I am of art I concluded that a psychiatrist, if a Librarian came to him with such a story of nocturnal subconscious caperings would, after lengthy probings of backgrounds and attitudes, come up with a diagnosis something like this.

He would say, my dear Fellow, it is clear that you are in a subconscious anxiety state about control, confinement, and utilization of the rising sea of print. While you consciously welcome the prospect of machine push button controls subconsciously you resist this. It is obvious that the castle is the citadel of traditional librarianship with all its drafty corridors and its numerous dark, musty, only partially controlled, poorly lighted and illy ventilated rooms of knowledge all bearing a mixture of room numbers compounded of the Library of Congress, the Dewey, the Expansive, the Bliss, the Ranganathan Colon and many lesser classification schemes. From what you have told me the corridors, and the rooms too, of this castle are replete with numerous directional signs in all kinds of languages. Go here, go there, look some place else, keep looking. Be sure not to miss the tenth and latest annex.

You, my friend, the psychiatrist would say, are the man on the bank. You know the castle is not the best possible castle but it is the only abode you know. It does serve after a fashion so you defend it. The damsel you rescue is Miss Electronic Controls, an exotic, tempermental, complex, confused, adolescent personality. She is full of promise and allure, but also of contradictions, and she speaks a weird gibberish which you, the defender of the citadel, do not understand and which she does not really understand either. You welcomed her eagerly hoping she would make your castle more livable. Then not understanding her and enraged by her gibberish you spurn her, hurling her back into the morass.

This then leaves you free to turn to your castle to supervise the building of the eleventh annex, to check the architect's plans for the twelfth and to begin preliminary planning for the thirteenth. Oh yes, and also to order up a few million more signs. Simple, dear chap, when one understands the subconscious.

WM. H. CARLSON
Director of Libraries
Oregon State University

BOOK REVIEWS

1/65–1R **Natural Language and the Computer.** 1963. Paul L. Garvin, ed. McGraw-Hill, New York. 398 pp. $8.95.

Mechanical translation had as its original aim the translation from one natural language to another natural language by means of processes carried on within a computer. Ordinarily, when we try to define such processes, we describe such activities as addition, subtraction, multiplication, or logical operations. A computer can transform a code on a punch card or piece of tape into another code which can activate a card-punch typewriter or printer. In such transformations, as in mathematical operations, formal relationships are established by the nature of the machine and the nature of the program between the input and output of the computer. In the case of natural languages, it is universally recognized that no finite set of formal transformations can translate one natural language as an input into one natural language as the output.

Within the past few years, this recognition has led to a gradual reduction in the amount of research effort in the field of machine translation. The best that any serious student in this field now proposes is some sort of a mechanical aid to human translation or some sort of mechanized dictionary look-up and print-out as a substitute for translation.

This general recognition that machine translation from natural language to natural language is a blind alley, however, has had the unfortunate result of creating a group of people interested in linguistics or natural languages in relation to the computer with nothing much to do. Looking around, these men have discovered the field of information retrieval, and they are now attempting to develop methods whereby information stores can be constructed out of inputs of natural language. It is this development that makes the subject of natural language and the computer of interest to documentalists, IR specialists, and librarians.

Mr. Garvin, the editor of this volume, states this development quite succinctly:

> Concretely, language-data processing for linguistic analysis will primarily include automatic linguistic analysis or at least automatic aids or automatic preliminaries to linguistic analysis (machine translation). Language-data processing for information handling includes such fields as machine translation, information storage and retrieval (if based on natural language), automatic abstracting, certain intelligence applications, and the like. . . . The area of linguistic information processing can be divided into two major subareas: (1) machine translation; and (2) information retrieval, automatic abstracting, and related activities, all of which may be summarized under the heading of content processing (p. 109).

Mr. Garvin recognizes that the aims he has set forth for natural language processing have been seriously questioned, and he is fair enough to present the essence of these questions:

> It is also worth noting in this connection that important negative opinions have been voiced with regard to both aspects of language-data processing. N. Chomsky, whose approach to language was presented earlier by Stockwell in "The Transformational Model of Generative or Predictive Grammar," takes the strong position that a discovery procedure—that is, a fixed set of rules for the discovery of relevant elements—is not a realistic goal for a science such as linguistics. This implies, of course, that automatic linguistic analysis also is an unreasonable proposition. Y. Bar-Hillel, a well-known symbolic logician and a philosophical critic of language-data processing, takes an equally clear-cut position. In a survey of machine translation in the United States conducted on behalf of the Office of Naval Research, he made the well-known statement that fully automatic, high-quality machine translation is impossible. He has voiced a similarly negative view in regard to other aspects of practical language-data processing (pp. 110,111).

Having presented these negative positions, however, Mr. Garvin has only a weak response to them:

> Needless to say, in spite of the need for objective criticism and an awareness of the difficulties involved, a more positive attitude toward the field is a prerequisite for active participation in the research (p. 111).

The question as I see it is not whether we have a positive attitude towards a field of research, but whether or not such a positive attitude is intellectually justified. Certainly, someone who wishes to conduct research in phrenology and astrology had better believe in phrenology and astrology before he begins, but such a belief is no guarantee that he will achieve valid results. The objections raised by Chomsky, Bar-Hillel and the reviewer, in other places, to the attempt to formalize natural languages by computer are serious enough so that they cannot be brushed aside by faith and pious hopes.

In one sense, the whole volume can be regarded as an attempt to answer the negative conclusions reached by Chomsky and Bar-Hillel. It presents this answer in five parts and an overview. The flyleaf describes these parts as follows:

> Part I deals with some approaches devised by professional linguists to conceptualize and analyze natural languages. Part II deals with mathematics and computation; and Part III attempts to relate the properties of natural languages to the problem of processing data on a computer — from a linguistic, a logical, and a mathematical standpoint. Parts IV and V deal with the two areas of language data processing which, at the present time, have shown the most tangible results: machine translation and information retrieval. The concluding section provides an evaluative overview of the field.

The degree to which natural language data processing has shown tangible results in the fields of machine translation and information retrieval is, of course, the heart of the problem for readers of this journal; and until such tangible results are specified, the reviewer finds himself unable to do more than reiterate the conclusion which Dr. Garvin has presented from the work of Chomsky and Bar-Hillel. If a research aim is, in principle, impossible, it is difficult to see how it could be made to yield tangible results of any value.

Beyond this question of tangible results, the book contains discussions of the relation of mathematics to linguistics, and the use of logical and statistical techniques in the analysis of language. But, in the end, it is a discussion not of results, but of hopes. These hopes are best summarized in the concluding paragraph of the contribution by M. E. Maron, "A Logician's View of Language-data Processing":

> Machines of the future will be designed in radically different ways. They will not be based on present design techniques which use formal logic. We will be able to communicate with future computers more effectively than is now possible. They will learn from experience and eventually, if we are ingenious and diligent enough, they will be able to handle

natural language as we do — given the proper educational period. As we learn more about language and machines for handling language, we shall learn more about the whole problem of knowledge itself, and out of this type of research may grow a science of knowledge (p. 150).

Being a philosopher by training, the reviewer may have understanding, if not sympathy, for what Maron means by a "science of knowledge;" but it is doubtful that such a hope will be relevant to those information technologists concerned with the design, implementation, and operation of information systems.

MORTIMER TAUBE
Documentation, Inc.
Bethesda, Md.

1/65–2R **The Baseball Program: An Automatic Question-Answerer.** 1963. A. K. Wolf, C. S. Chomsky, and B. F. Green, Jr. Massachusetts Institute of Technology, Lincoln Laboratory, (Vol. 1, Lincoln Laboratory Technical Report 306), Lexington, Mass., April 1963. 201 pp.

The Barnstable Summer Study Group, convened in 1958 to study military communications systems, recommended that experiments be conducted on the programming of a computer to provide answers to questions put to it in natural English language. An obvious first step would be to write a computer program that would answer a limited variety of questions about a limited subject matter. Wolf and his associates at the Lincoln Laboratory have produced just such a program, designed to answer natural language questions of relatively simple structure and content on the subject of baseball. Since its vocabulary is reasonably small and its data fairly regular, this topic was regarded as particularly suitable for experimental purposes.

The Baseball Program, written in the list-processing language IPL-V, consists of two major parts, a Language Analyzer and a Data Processor. The data, stored in the form of tree structures, consist of the month, day, place, team and scores for games played in the American League in 1959. The program will handle demands for information explicitly present (such as "What was the Yankees' score on May 10?") and certain demands for information implicitly present (for example, "How many times did each team win at home before July 4?").

A question is read into the computer from punched cards, and the words and idioms are looked up in a dictionary. This dictionary contains all the words and idioms (e.g. how often, municipal stadium, took place) recognized by the Baseball Program. The dictionary definition gives information on part of speech, meaning, etc., needed in the Syntactic Analysis and the Content Analysis. The syntactic analysis used in Baseball, which is minimal, is modeled after the work of Zellig Harris in the automatic bracketing of phrases in English sentences. Specifically, this routine brackets the phrases in the question, decides whether the verb is active or passive, locates the subject and object, and determines whether the question requires either a yes-or-no answer or a content answer. All this information is required for the final operation of the Language Analyzer, the Content Analysis.

The Content Analysis provides input to the data processing routines, converting the natural language question into a "spec list," a formalized representation of content which is acceptable to the Data Processor. For the simple question mentioned above, the spec list would be as follows: (TEAM = YANKEES, SCORE = ?), MONTH = MAY, DAY = 10. The specification list formed for a question is used by the search routines to extract relevant information from the data and to organize the results for further processing where needed. The Responder prints out the answer produced by the Processor.

The present report explains in some detail the procedures and algorithms used in each section of the program. For the non-programmer, it is not easily digestible. This pioneer work of Wolf and his colleagues, however, must be regarded as essential reading for any documentalist or far-sighted librarian who believes that, ultimately, an "information retrieval system" should inform instead of merely spewing out lists of document numbers.

F. WILFRED LANCASTER
Herner and Company
Washington, D. C.

1/65–3R **A Contribution to the Early History of Documentation (Adatok és felismerések a szakirodalmi dokumentació tortenetenek kezdeteihez).** 1964. Dr. Ivan Polzovics. Országos Müszaki Könyvtár es Dokumentációs Központ, Budapest. 80 pp.

Ivan Polzovics' eighty-page booklet is more voluminous than all other works on the early history of documentation combined. Many of the available historical accounts are modestly hidden in prefaces and introductions to textbooks and treatises, usually with an apologetic reference to the obscurity of the genesis and infancy of documentation. Among the pioneers who drew the first sketchy maps of this hazy realm, S. C. Bradford, P. Otlet, Th. P. Loosjes, J. H. Shera and M. E. Egan should be mentioned.

There are at least three common approaches in attempts to identify the roots of documentation. The most frequently encountered is based on the truism that knowledge in any form is informative and classifiable. Therefore, it is only a matter of degree in determining at which stage of the evolutionary scale documentation emerges as a distinct branch of knowledge. Hence, the prelude to the discipline may be recognized in primitive bibliographic compilations (e.g., the classified references in the appendix of the Venerable Bede's Ecclesiastical History of Britain, A.D. 731) or in the ingenious first abstract journals of the seventeenth century. Since this is virgin territory for the historian of science, every justifiable approach merits praise.

It has also been claimed that documentation first became visible as a link between librarianship and scientific activity. This book-oriented approach is suggested mainly by the historians of libraries who, with great pride, point to the specified subject-grouping of Assurbanipal's clay tablets in Nineveh.

Polzovics' Early History carries no reference to the Stone (or Clay) Age of documentation. He merely scrutinizes those patterns which, having no connection with librarianship, mark the original forms, the novel approaches, the genuine methods of the newly evolved discipline. January 5, 1965, will be the 300th anniversary of the first abstracting journal in France, an event which should be regarded as the foundation of modern scientific information. "Le Journal des Sçavans," a periodical for learned news and information, was created with the sole aim of abstracting scientific and literary books of contemporary Europe, reviewing new inventions and discoveries, mainly in the realm of physics, chemistry, astronomy and anatomy.

The author analyzes the significance of "Le Journal des Sçavans," whose weekly publication constituted an interdisciplinary forum, a stimulating organ for the exchange of ideas and, primarily, an evaluative record of scientific accomplishments from the later seventeenth century onwards.

It should be mentioned that Polzovics does not deserve the sole credit for demonstrating the prestige of "Le Journal des Sçavans." This was already well illustrated by many European and American writers. Indeed, a much more complete historical account is to be found in David A. Kronick's "A History of Scientific and Technical Periodicals" (Scarecrow Press, New York, 1962).

Polzovics' merit lies elsewhere. Based on demonstrative proofs, he claims the founder of scientific documentation to be no less a person than Jean Baptiste Colbert, the guiding spirit of French mercantilism. The story of these investigations makes lively reading, is full of surprise turns and, incidentally, sheds new light on the birth of "Le Journal des Sçavans."

Among the vast collection of Colbert's correspondence, Polzovics found a letter from Jean Chapelain, poet laureate and a founding member of the Académie Française, to Colbert who as chief minister of Louis XIV was an ardent proponent of industrial growth and a powerful promoter of the arts and sciences. Referring to a recent conversation, Chapelain warmly recommends a councillor of the Law Courts of Paris named Denis de Sallo for the post Colbert has in mind. According to Chapelain, this candidate is a versatile man who, among other talents, can "stimulate the awkward scientists to read and to prepare extracts of their readings for him." The ideal editor for an abstract journal! Eventually, when "Le Journal des Sçavans" appears two years later, the same Denis de Sallo becomes its first editor in chief. From this and from other supporting facts, Polzovics makes it

obvious that Colbert carried out a painstaking search for a qualified man who would be proficient in languages and able to organize the abstracting service of the journal of which he himself was founder.

Three months after the first issue of "Le Journal des Sçavans," Denis de Sallo quit the editorial chair, addressing his letter of resignation to Colbert who, apparently, was the sole authority in the affairs of the journal. In his letter, de Sallo bitterly criticized his co-workers for their poor contributions and their unreliability. This editorial crisis, to which politico-religious factors also contributed, caused the journal to cease publication for over a year.

The next appointee was the cleric Jean Galois, an expert in physics, mathematics, astronomy and linguistics; a fortunate choice to head an abstracting journal. Obviously, Colbert wanted to be certain that, after such an abortive start, the periodical would now function smoothly in the execution of its task. Galois was Colbert's own choice; the tutor of his children, the abbot also lectured the great minister himself on Latin and Greek. Born the son of a draper, Colbert badly needed some improvement of his insufficient education and Galois became not only his longstanding instructor but also a permanent traveling companion and trusted friend. His appointment clearly reflects the importance Colbert saw in the realization of his concept of the abstracting journal.

The innovation whose success was thus furthered by Colbert's perseverance, created a model of ready-to-use scientific information, badly needed in a rapidly changing world of scientific revolution. The content of the journal was soon to be translated and published simultaneously in Germany and Italy. Two months after its first appearance, Henry Oldenburg, Secretary of the British Royal Society, introduced his "Philosophical Transactions: Giving Some Accompt of the Present Undertakings, Studies, and Labours of the Ingenious in Many Considerable Parts of the World." In order to demonstrate the exemplary value of the French initiative, Polzovics attempts to prove that "Le Journal des Sçavans" was the springboard for modern information methods. Its success led not only to the foundation of numerous abstract and review journals, but also triggered the establishment and growth of periodicals for full-length papers in all the sciences.

As an advantageous "side effect," a new type of scientific activity evolved, that of the abstracter, whose business was to evaluate the merits of each publication in an unbiased way and to summarize its essentials to promote the dissemination of new scientific knowledge.

In the postscript the author suggests that the 300th anniversary of the publication of the first abstracting journal represent a notable historical milestone in the development of scientific information. National and international documentation organizations might find this date a suitable opportunity to recall the historical roots of documentation and to pay tribute to the memory of is innovator.

GEORGE EMBER
Documentation Service
Institute of Experimental Medicine
University of Montreal

LITERATURE NOTES

CHARLES P. BOURNE, Editor

Contributions of abstracts from readers and suggestions of books and articles for review or inclusion in this bibliography will be welcome and are actively solicited. Volunteer abstracters and reviewers are needed, as well as people with special linguistic talents. All copies of reprints, reports, and correspondence for this section should be addressed to Mr. Charles P. Bourne, Stanford Research Institute, Menlo Park, California. In order to increase its coverage, this section will include some abstracts copied or adapted from other publications. The American Documentation Institute is not able to supply copies of the publications abstracted or cited.

Abstracters

The following persons have prepared some of the abstracts for this issue. Their names are listed here to express the editor's thanks and appreciation as well as to explain the initials that were used to sign the abstracts.

T.G.F.	Trudy G. Fisher
F.M.	Fenster Moop
D.C.W.	David C. Weber
C.D.G.	C. D. Gull
V.E.Y.	Virginia E. Yagello
J.A.W.	Judith A. Werdel
P.M.B.	Peter M. Bernays
J.J.C.	J. J. Cirrito
P.A.R.	Phyllis A. Richmond
B.K.D.	Bernard K. Dennis

Contents

Starting with this issue, separate copies of this section of *American Documentation* may be obtained as reprints on a subscription basis for $4.00 per year. Requests should be directed to ADI, 2000 P Street, North West, Washington, D. C., 20036. It is hoped that in the very near future it will be possible to provide Literature Notes as a publication separate from *American Documentation* and with increased coverage. The availability of separate reprint sections is a start toward that objective.

Index of Authors and Corporate Sources Referred to in this Issue

(The numbers refer to the terminal digits of the abstract serial numbers in this issue. The letter *R* refers to Reviews.)

Journal Abbreviations

Starting with this issue, many of the journal titles will be represented by their corresponding CODEN representation as defined in the two publications, Coden for Periodical Titles. American Society for Testing and Materials, Phila., Pa. 1963. ASTM Special Technical Publication No. 329, and Coden for Periodical Titles. Supplement 1. 1964. ASTM Special Technical Publication No. 329–S31. The CODENS are being used in place of the full titles primarily to reduce typing and typesetting work, and to provide a more compact representation of the citation. The following table will help translate the CODENS for this issue.

ALBL	ALA Bulletin
AMDO	American Documentation
ASLP	Aslib Proceedings
BATR	Battelle Technical Review
BEHS	Behavioral Sciences
BHOR	Business Horizons
BUAT	Business Automation
CACM	ACM Communications
CENG	Control Engineering
CPGR	ACM Computing Reviews
DKMN	Dokumentation
DTMN	Datamation
IBMJ	IBM Journal
IEEW	IEEE Trans. on Eng. Writing & Speech
IPAL	Inland Printer and American Lithographer
IREL	IEEE Trans. on Electronic Computers
IRTE	IEEE Trans. on Eng. Management
ISSR	Information Storage & Retrieval
JACM	ACM Journal
JCHD	J. Chemical Documentation
JDOC	J. Documentation
LCIB	Library of Congress Information Bulletin
LIBJ	Library Journal
LIBL	Library Literature
LIBQ	Library Quarterly
LIBT	Library Trends
LRTS	Library Resources & Technical Services
NADO	Nachrichten für Dokumentation
NAMN	National Micro-News
REDO	Revue Internationale de la Documentation
RPRV	Reproduction Review
SCIE	Science
SPLB	Special Libraries
UNDL	UNESCO Bull. for Libraries
WLDU	Wilson Library Bull.

Published Reference Works

1/65–1 **Supplement B to the Guide to Microreproduction Equipment (1962).** April 1964. Hubbard W. Ballou, Ed. National Microfilm Assoc., Annapolis, Md. 136 pp., paper bound; price (non-members) $5.00. Member rate $3.00.

The Supplement continues the up-dating of the 2nd edition of the Guide . . . by noting changes, corrections and new equipment appearing between April 1963 and March 1964. There are 55 new items listed, and major revisions in 13 other models. Minor revisions are reported for another 50 items, and 18 pieces of equipment have been deleted from previous listing. To make Supplement A and Supplement B easier to use, an index to these two supplements has been included. A new listing of Manufacturers and Distributors has been prepared, noting the addition of 29 companies over the 2nd edition, and a dropping of 6 companies.

1/65–2 **Glossary of Terms for Microphotography and Reproductions Made from Micro-Images.** 3d. ed. 1964. D. M. Avedon, Ed. Monograph No. 2 of the National Microfilm Assoc., Annapolis, Md.; price (non-members) $3.50. Member rate $2.50.

1/65–3 **A Complete List of Periodicals Evaluated by Documentation Centers of East Germany. Technology, Natural Sciences and Related Fields.** Institute for Documentation of the German Academy of Sciences, Berlin. 2nd. rev. and enl. ed. Leipzig, Verl. fuer Buch und Bibliothekswesen, 1963. 528 pp. (Condensed from a review by W. Dux, DKMN 1964, No. 1, pp. 28–29.) In German.

A total of 6,300 titles in this second edition reflects the expanded activity of the documentation network in East Germany and the increasing number of new periodicals in the field of technology and natural sciences. The main section gives reference

data on each periodical, including code numbers of the evaluating documentation center and call numbers of the source library. An index section lists all the titles arranged by code numbers. The next section of the book is a list of all the documentation centers and libraries which participated in this project. Since the documentation centers process the majority of all the important publications of the whole world, the catalogue represents a selected international bibliography of technical and scientific periodicals of this branch. The list of periodicals in library holdings is not complete, however, since a number of libraries, some of which are known to stock many periodicals, were not included in the survey. (Doc. Inc.)

Information Resources — General

1/65–4 **Scope and Duplication of Coverage in Two National Information-Processing Systems.** 1963. V. Slamecka. *Automation and Scientific Communication,* pp. 171–172 (see 7/64–151).

1/65–5 **The Materials Information Network.** 1963. Edward Dugger. *Automation and Scientific Communication,* pp. 217–218. (see 7/64–151).

1/65–6 **The Role of the Office of Technical Services in the Interchange of Documents.** 1963. Lillian A. Hamrick. *Automation and Scientific Communication,* pp. 219–220 (see 7/64–151).

1/65–7 **A Program for Dissemination of Specific Data on Materials.** 1963. H. Thayne Johnson. *Automation and Scientific Communication,* pp. 295–296 (see 7/64–151).

1/65–8 **Role of "Letters" Journals in Primary Distribution of Information.** 1963. George L. Trigg. *Automation and Scientific Communication,* p. 319 (see 7/64–151).

"Letter" journals, which developed out of the "Letters to the Editor" section of research journals, provide rapid dissemination of results which are judged likely to have marked effects on the work of a substantial number of people. This is accomplished by keeping the communications brief, reviewing them promptly, and making use of rapid publication methods (author).

1/65–9 **A National Plan for Science Abstracting and Indexing Services.** 1963. Raymond A. Jensen. *Automation and Scientific Communication,* pp. 339–342 (see 7/64–151).

1/65–10 **The Aerospace Research Applications Center** (Indiana University). Summer 1964. Arthur M. Weimer and Howard L. Timms. *Business Horizons,* 7(2): 93–100.

1/65–11 **Foreign Literature of Chemistry.** 10 May 1963. J. L. Wood, K. L. Coe, and G. O. Platan. SCIE 140(3567): 610–613.

Case Studies of Manual Systems in Operation

1/65–12 **Indexing Costs for 10,000 Documents.** 1963. L. H. Linder. *Automation and Scientific Communication,* pp. 147–148 (see 7/64–151).

Detailed costs for equipment, supplies, indexing and input to a machine document address storage system are presented for a collection of 10,000 items. For each document an average of 12.6 access points, including 9.2 subject approaches, was provided. The effort involved approximately 4 man-years of work, and the total cost of the system was less than $30,000. This favorable figure was possible because an adequate ready-made thesaurus of indexing terms was available and because the "peek-a-boo" type equipment used was much less expensive than most other devices offering comparable speed of operation and search logic possibilities.

In June 1961, the Aeronutronic Division acquired a Termatrex 10-track model (now called the Jonker 202), a Termatrex reader, a supply of "Radex" cards, and the necessary card holders for the purpose of storing and retrieving references to a portion of the external report literature being received by the Division. Since then, more than 10,000 scientific and technical reports have been analyzed, indexed and processed with the aid of this system. Following a description of the equipment and the method, this paper reports detailed costs associated with indexing of the first 10,000 reports (author).

1/65–13 **An Operational Information Retrieval System in the Field of Cryogenics.** 1963. Neil A. Olien. *Automation and Scientific Communication,* pp. 157–158 (see 7/64–151).

1/65–14 **Radiation Shielding Information Center Information Retrieval System.** 1963. S. K. Penny, D. K. Trubey, and M. B. Emmett. *Automation and Scientific Communication,* pp. 251–252 (see 7/64–151).

1/65–15 **Experiences With Peek-A-Boo Cards.** 1964. J. Toman. *Technicka Knihovna* (Technical Library), Prague, No. 6, pp. 186–191. In Czech.

The author participated in a research project which included the testing of visual punch cards, handled manually, for purposes of filing and information retrieval in the building industry. The inverse system of indexing used for the reference file, the search mechanism based on positional concidence of perforations, the problems of open-endedness and provisions for additions and modifications, capacity and density of punching, multidimensionality and rapidity of search, and other related topics are discussed and analyzed. The most important but also most difficult part of the peek-a-boo system is setting up a good reference file. Current Aritma cards produced in Czechoslovakia can be easily provided with a 999 hole punch grid. The business machine corporation, Kancelarske Stroje, can also furnish cards with 1,000, 2,000 or 4,000 punching positions. The U.S. Microcite system has not yet been introduced in Czechoslovakia. The author states in conclusion that the peek-a-boo system has great advantages and should be used more extensively in the future. (T.G.F.).

1/65–16 **Screening for Best Qualified Officers.** 1964. *Army Information Digest* 19(8): 30–31.

Describes the use of Termatrex, an optical concidence (peek-a-boo) system, at the Army Officer Personnel Directorate. The information retrieval system stores data on experience, education, and special qualifications of approximately 90,000 officers. Its use permits selection of properly qualified officers, on a world-wide requirement basis, more rapidly than through a previous technique. (J.J.C.).

1/65–17 **Edge-notched Cards in Documentation Practice.** 1964. H. G. Krey, DKMN, No. 2, pp. 39–43. In German.

Experiences with a punch card file on plant protecting agents and pesticides are reported. If the expected quantity of entries is relatively great, provisions should be made for eliminating outdated material. Sometimes a single reference item requires several cards since it is not possible to notch several code numbers in one punching field. The average selection, preparation, and storage time per reference item for such a system is twenty-one minutes and does not include abstracting. The average cost of a search for bibliographic references utilizing the punch card system is 56.00 marks as compared with the estimated cost of 600.00 marks if the search is done by a scientist without the file. Consideration is given to changing the system to a machine punch card file. (Doc. Inc.).

Rapid and accurate preparation of large amounts of alphabetic information as input for computer searching is feasible using flexowriters. By this method, the University of Pittsburgh Health Law Center prepared 4,800,000 lines of legal text, mostly consisting of the statutes of various states and the federal government (author).

1/65–49 **Practical Operating Experience with a Tape Controlled Photon S-560 Unit.** 1963. D. A. Luce, M. P. Barnett. *Automation and Scientific Communication,* pp. 293–294 (see 7/64–151).

A Photon S-560 paper tape drive photo composing unit has been in use in the Cooperative Computing Laboratory, Massachusetts Institute of Technology, since April, 1963. The Flexowriter-computer-S-560 System can be used to set material with typographic quality (author).

1/65–50 **Photographic and Computer Systems in Biomedical Information Handling.** 1963. Lawrence C. Kingsland, Jr. *Automation and Scientific Communication,* pp. 313–314 (see 7/64–151).

Both photographic and computer-based systems are proposed in attaining more comprehensive understanding of specialized biomedical terms and concepts (author).

1/65–51 **Microfilm and Hard Copy Transmission from Document Retrieval Centers.** 1963. G. F. Stafford. *Automation and Scientific Communication,* pp. 333–334 (see 7/64–151).

1/65–52 **Automating the Serial Record.** June 1964. Joseph Becker. ALBL pp. 557–560.

1/65–53 **Computer Generation of Photocomposing Control Tapes. Part 2. The PC6 System.** April 1964. M. P. Barnett, D. J. Moss, and D. A. Luce. AMDO 15(2): 115–120.

1/65–54 **A New Era in Library Technology.** June 1964. Hans A. Lustig. BUAT p. 37.

A description is given of the use of sequential card composition for repetitive publications such as library catalogs.

1/65–55 **Use of the Friden Flexowriter in the Library of the Atomic Energy Research Establishment, Harwell.** March 1964. C. W. J. Wilson. JDOC 20(1): 16–24 illus.

To reduce the typing duplication between acquisition lists and catalog cards and reproduce cards for 15 branch libraries, a Flexowriter Type FPC 8 Recorder/Reproducer is used with edge-punched cards and self-adhesive labels for identifying each title. It is planned to add the carbon acetate ribbon attachment. The typist copies directly from the document with the help of a standard worksheet prepared by a cataloger. Cards were chosen instead of tape because of greater ease of handling and filing; yet the stock is expensive, especially since a typical entry uses 4–6 cards plus 1–2 labels and corrections require retyping the card. Disadvantages are the noise and the fact that four typists must be trained to the instrument to achieve economic operation by reasonably full utilization of the working day. With the entries used for several lists, substantial savings in man hours have resulted (D.C.W.).

1/65–56 **Review of Copying Methods: 1963.** Spring 1964. Peter Scott. LRTS 8(2): 131–144.

Notable were the first International Congress on Reprography, the "measles" developed by microfilm, an increase in photocopying in lieu of loan, and a 25% increase in the number of titles available in microtext. A detailed review is also given of new equipment—including the Miracode and Radir systems—the emergence of microfiche, the programs of the Library Technology project, and significant publications (D.C.W.).

1/65–57 **One Picture Is Worth a Thousand (Typed) Words.** Spring 1964. William E. Jorgensen. LRTS 8(2): 196–198. Illus.

Catalog card reproduction on a Xerox 914 is achieved at a rate of about 200 per hour including cutting, punching, and assembling sets. Aluminum weights hold the six originals; copies are on .007 inch thick card sheets, although the Xerox Company does not recommend the use of paper above .006. Operational details are given (D.C.W.).

1/65–58 **Combined Procedures for Technical Processes.** Summer 1964. Earl Farley. LRTS 8(3): 257–265.

Use of photography, xerography, and offset printing is described in technical processes at the University of Kansas Libraries. A microfilm camera has been used to obtain Copyflo prints for decision making during book selection, for purchase claims, for "brieflisted" catalog cards, and for card duplication. Masks are used for several of these applications. Catalog cards are now duplicated by transferring copy four up on a Xerox 914 to a pre-perforated 8½" x 12" multilith offset master. Some operating details and costs are included (D.C.W.).

1/65–59 **Equipment and Methods in Catalog Card Reproduction.** Summer 1964. Joseph H. Treyz. LRTS 8(3): 267–278.

This revision of a paper delivered in June 1962 and published in *Library Furniture and Equipment* is a review of problems and methods of reproduction. It covers printed cards, manual and automatic typewriters, stencil duplicators both postcard size and full size, offset presses, xerography, and dye transfer. The conclusion is that a small library should buy a postcard size stencil duplicator; a medium size library should have a full size stencil duplicator; and an offset machine, with a camera to make masters if needed, is suited to the larger libraries. Further developments are anticipated in processes and machine applications (D.C.W.).

1/65–60 **Library Card Reproduction by Xerox Copyflo.** Summer 1964. Allen B. Veaner & John Fraser. LRTS 8(3): 279–284.

The Harvard College Library changed from offset printing to microfilm and Copyflo printing in the summer of 1963. The background and preliminary analysis are discussed together with the system techniques, advantages, disadvantages, and costs. It is anticipated that the new techniques will have a lifetime of three to five years, after which further advances may be expected from the installation of computers (D.C.W.).

1/65–61 **Computer Typesetting: An Evaluation of the Problems.** Dec. 1963. C. J. Duncan, et al. *Printing Technology.* Pp. 133–143.

The paper reviews the problems associated with the use of digital computers for editorial work, allocation of composing machines and automatic text preparation. The particular problems of hyphenation and page make-up are considered in detail. The progress of work on optimizing the routines for a limited number of cases is discussed, with examples of the experimental work so far produced at Newcastle and elsewhere, and with indications of the direction of future work (authors).

1/65–62 **How a Computation Center Could Help to Solve Documentation Problems.** Jan.-Feb. 1964. Andre Deweze. UNDL 18(1): 1–12.

According to the author's calculations, the additional expenditure required to automate documentation in a large industrial firm equals the cost of two senior engineers. It covers the rental of computing equipment, consumable supplies, and salaries of additional staff, microfilm laboratory operatives, an assistant documentalist and an analyst. The additional expenditure would

or indexing and processing library material. The ARITMA T 320 omputer was used to index and classify periodicals according to arious methods and viewpoints. The machine-sorted punch cards ere useful for producing catalogs for general or specific purposes. pecial attention was given to the cataloging of foreign language material, coding such material according to subject matter, language, country of origin, etc. Tests were not limited to the ARITMA T 320, but covered the entire line of ARITMA machines nd the application of digital computers in libraries (T.G.F.).

Application of Equipment to Library Operations

/65–39 **The Economics of Book Catalog Production. A Study Prepared for Stanford University Libraries and the Council on Library Resources.** 31 May, 1964. R. M. Hayes and R. M. Shoffner. [Sherman Oaks, California, Advanced Information Systems Division, Hughes Dynamics, Inc.] 110 pp. Processed. Obtainable from: Office of the Director, The Stanford University Libraries, Stanford, California. Xerox copies unbound, $10; 35 mm. positive microfilm, $5; also available on interlibrary loan. Not available from the Council on Library Resources.

This report was intended to assist Stanford in the economic and procedural questions involved in deciding what form the catalog for a new undergraduate library should take. It is also expected to assist other libraries in making such a decision. The new Stanford library is expected to begin with some 40,000 titles (60,000 volumes) and, for the sake of this study, to add 10,000 titles per year until it totals 100,000 volumes. The possibility of a catalog, which, in addition to being kept in each pavilion on four floors of the library, could be distributed about the campus and perhaps available in subject parts for student purchase is under consideration.

Six basic methods for producing catalogs in book form are considered: 1. reproduction of Library of Congress cards by photo offset from shingled copy; 2. reproduction from cards specially typed and shingled; 3. reproduction from typed pages; 4. reproduction from tab cards; 5. reproduction from copy prepared for sequential cameras; and 6. reproduction by means of several forms of computer composition. Alternatives for several methods are also noted. Time and cost factors are analyzed and information is also expressed in algebraic form.

In the introductory section, the general nature of the approach adopted is described. General formulas for estimating costs for various specific manual and mechanized methods are given in the second section and sample pages from book catalogs produced by these methods are shown. As an illustration of the application of the general approach, details are given in the third section of the comparative economics for the various methods in the instance of the Stanford library. The appendices contain a discussion of characteristics of applicable equipment, the derivation of functions for seven specific operations, and sample statistics for characters in L.C. cards, for rate of manual filing, and for Stanford's card duplication. There are also a short annotated bibliography and a list of equipment manufacturers used in the study.

1/65–40 **A System for Transcribing Printed Text into a Machine Readable Format.** 1963. John L. Bennett. *Automation and Scientific Communication*, pp. 141–142 (see 7/64–151).

The printed text from 50 articles on speech analysis and synthesis has been transcribed into a computer-readable format. This "short" summarizes the steps in the copying method and serves as a basis for exchange of experience and of ideas on alternate transcription methods (author).

1/65–41 **Document Reproduction from Microfilm Is Important in Speed and Cost by Procedure Making Six Copies Per Shot.** 1963. Walter H. Rupp. *Automation and Scientific Communication*, pp. 151–152 (see 7/64–151).

The document microfilming and copying procedure described here demonstrates a practical system for increasing the speed of producing hard copies by six-fold and reducing the cost of copy to about one-quarter normal. The procedure can improve the effectiveness of moderate-sized IS&R systems (author).

1/65–42 **Satire Display Subsystem.** 1963. John P. Roach, Jr. *Automation and Scientific Communication*, pp. 165–166 (see 7/64–151).

This document proposes new equipment which will enable "self service," "hard copy" information retrieval (author).

1/65–43 **The Use of a Cathode Ray Tube Display Console for Editing Textual Information.** 1963. Lawrence F. Buckland. *Automation and Scientific Communication*, pp. 179–180 (see 7/65–151).

Experiments are being performed to test the utility of a computer-connected cathode ray tube display in the man-machine processes of text editing and indexing, where a final output must be in machine as well as visual form. Results so far show that editing can be performed faster than with a typewriter console, and that the immediate display response is very useful when several sequential editing steps are needed in a local area of text (author).

1/65–44 **State-of-the-Art: Remote Interrogation of Stored Documentary Material.** 1963. Herbert Ohlman. *Automation and Scientific Communication*, pp. 193–194 (see 7/64–151).

Remote interrogation is defined. Remote manipulation techniques for material in human readable form are compared with those for machine-manipulatable media. Analog and digital transmission means, the hard-copy dilemma, and interrogation modes are discussed, and the input problem is highlighted (author).

1/65–45 **Some Observations on Mechanization of Library Processes.** 1963. E. M. McCormick. *Automation and Scientific Communication*, p. 195 (see 7/64–151).

1/65–46 **Programming and Coding Techniques Used in MEDLARS.** 1963. Anne C. Davis, Frank E. Booth. *Automation and Scientific Communication*, pp. 223–224 (see 7/64–151).

This paper describes some of the programming logic of the MEDLARS system. The Demand Search Module retrieves citations from the citation file initially on the basis of "significant" elements; i.e., elements having the lowest frequency of usage which must be present for a citation to satisfy the request. The resultant file which is reduced in size is then passed against a decision table for the final selection of citations. The Report Generator Module formats the citation of printer paper, cards, or photographic film. The Output Processing Module formats *Index Medicus* and recurring bibliographies, which are automatically updated by the system. The final output is processed by the Graphic Arts Composing Equipment for later printing (author).

1/65–47 **Analysis of the Role of the Computer in the Reproduction and Distribution of Scientific Papers.** 1963. Joseph H. Kuney. *Automation and Scientific Communication*, pp. 249–250 (see 7/64–151).

The American Chemical Society has begun an analysis of the role of the computer in related aspects of the reproduction, distribution, and retrieval of scientific information. Initial work will attempt to solve problems of photocomposition via computer (author).

1/65–48 **Use of Flexowriters to Prepare Large Amounts of Alphabetic Legal Data for Computer Retrieval.** 1963. John F. Horty, Thomas B. Walsh. *Automation and Scientific Communication*, pp. 259–260 (see 7/64–151).

permit indexing 20,000 instead of 10,000 documents per year with facility for rapid retrieval of the information they contain, the mechanical production of a cumulative index of the keywords in the bibliographical bulletin above, the possibility of publishing a KWIC index, quick replies to requests for bibliographical information, the publication of a review of reviews in fifty copies, and a system for the automatic exploitation of microfilms of all the documents indexed (C.D.G.).

Application of Data Processing Equipment to Classification, Indexing, and Text Processing

1/65–63 **Automatic Processing of Metallurgical Abstracts for the Purpose of Information Retrieval.** Feb. 1964. Jessica S. Melton, Pamela W. Reeves, Celeste F. Hespen. Center for Documentation and Communication Research, School of Library Science, Western Reserve Univ., Cleveland, Ohio. 104 pp. Interim report NSF–2.

The aim of this project is to develop the test methods for the automated processing of previously generated conventional abstracts for information retrieval. The report describes progress in an attempt to automate subject indexing, which is perhaps the most difficult and expensive task of information systems when performed manually. Abstracts from the metallurgical section of *Chemical Abstracts* are punched on Flexowriter tape and converted to magnetic tape. The typographical characteristics of the printed text are preserved. Computer procedures are designed for processing the abstracts on a level above straight dictionary look-up while avoiding total linguistic analysis. Metallurgical terms in the text are located and grouped according to a strict subject-indexing rationale.

Experience is reported in the analysis of 223 abstracts comprising 1290 sentences and approximately 20,000 words. From this corpus, a lexicon of 3200 words has been compiled and classified, and 1500 rules for grouping metallurgical terms into machine-searchable index structures have been formulated. Plans for testing the system and preliminary work in this direction are described (authors).

1/65–64 **Toward the Establishment of a Computer Based Classification System for Scientific Documentation.** Feb. 19, 1964. Harold Borko and Myrna D. Bernick. System Development Corp., Santa Monica, California. 47 pp. Reports TM–1763 and AD–433 830.

Experiments were conducted to determine whether factor-analytically derived classification categories are reliable, reasonable, and useful. A sample of 997 abstracts of psychological research papers was selected and divided into an Experimental and a Validation Group. These were factor-analyzed separately and the resulting categories compared to each other and to a previous study. The results show that factor-analytically derived classification categories are reliable and reasonable. Experiments were also performed to compare manual and automatic document classification. A contingency coefficient was used to measure inter-rater consistency of classification and the similarity of manual and machine classification. It was concluded that certain specialized document collections could advantageously use a factor-analytically derived classification schedule and automatic document classification (authors).

1/65–65 **Automatic Subject Indexing from Textual Condensations.** 1963. V. Slamecka, P. Zunde. *Automation and Scientific Communication,* pp. 139–140 (see 7/64–151).

Since methods of automatic abstracting and indexing have not yet led to an acceptable product, it is appropriate to inquire whether the intellectual effort in the subject analysis of documents by humans can be exploited more economically. Such a possibility exists, of course, whenever human effort can be formalized or when a formal matrix is available against which this effort can be interpreted, as is manifest from such practices as the compilation of subject bibliographies from catalog or punched cards without the re-examination of document subjects. The present paper describes a method of automatic preparation of conventional printed and machine-stored subject indexes from textual condensation, employing as the interpretative matrix an authoritative vocabulary of indexing (searching) terms (authors).

1/65–66 **Producing Indexing Aids Such as a Dictionary or Thesaurus from a Single Punched Card Deck.** 1963. J. F. Tinker, R. B. Smith. *Automation and Scientific Communication,* pp. 173–174 (see 7/65–151).

By use of relatively simple equipment, such as a sorter, and tabulator or document writer, a dictionary or thesaurus can be prepared. The format of the thesaurus can be designed to display a variety of relationships between indexing terms. It can be varied without changing its intellectual content to suit the habits and viewpoints of the user. Corrections, additions, and reverse entries to these indexing and retrieval aids can be made simply and easily. The deck can be used in conjunction with the Listomatic method or computer-controlled photo-composing devices to produce high-quality print-outs (authors).

1/65–67 **Automatic Retrieval and Selective Dissemination of References from Chemical Titles: Improving the Selection Process.** 1963. Robert R. Freeman. *Automation and Scientific Communication,* pp. 213–214 (see 7/64–151).

An IBM 1401 computer program for retrieval of references from *Chemical Titles* has been developed. Future areas of research include selection based, in addition to keywords and initially-occurring stems, on author names, journal reference codes, non-initial morphemes and stems, and both subjectively and statistically determined relevance (author).

1/65–68 **A Progress Report on an Experiment in Semiautomatic Indexing Conducted by the AEC Division of Technical Information Extension.** 1963. John Sherrod. *Automation and Scientific Communication,* p. 215 (see 7/64–151).

Encoded paper tapes generated concurrently with the typing of abstracts have been used to generate various indexes to the abstracts. In preparing the subject index, main subject headings selected manually from the title and abstract are arranged by a computer so that the title, often in modified form, follows each of the main entries. The retrieval efficiency of this index has not been compared objectively with conventional indexes (author).

1/65–69 **An Experiment in Automatic Indexing of French Language Documents.** 1963. F. Levery. *Automation and Scientific Communication,* pp. 235–236 (see 7/64–151).

1/65–70 **Computerization of Book Catalogues at the Columbia, Harvard, and Yale Medical Libraries.** 1963. Frederick G. Kilgour, Ralph T. Esterquest, Thomas P. Flening. *Automation and Scientific Communication,* pp. 299–300 (see 7/64–151).

1/65–71 **Characteristics of Programs for KWIC and Other Computer-Produced Indexes.** 1963. W. W. Youden. *Automation and Scientific Communication,* pp. 331–332 (see 7/64–151).

This paper defines and discusses most of the different types of computer-produced indexes, including keyword, author, corporate author, citation and conventional subject-heading indexes; also, the computer input and output formats for indexes, double look-up, coden and a look at the future.

Computers have been used to produce many different types of indexes; most have been "current awareness" indexes, although a few cumulative or retrospective indexes have also been produced. This paper discusses computer input and output formats, most of the different types of computer-produced indexes, problems of double look-up, codens and a look at the future (author).

1/65–72 · **Compilation and Computer Updating of a Medical Sciences Thesaurus.** 1963. Lynda Lee McGee, William J. Holliman, Jr., Allan Z. Loren, and Gordon D. Adams. *Automation and Scientific Communication*, pp. 347–348 (see 7/64–151).

A "Medical and Health Related Sciences Thesaurus" has been compiled as an indexing guide and entered on computer for updating and periodic print-outs. The first printed edition of the thesaurus contains approximately 12,200 biomedical terms, of which 6,700 are main headings; 2,200 subheadings; 2,300 "see" references; and 1,000 instructional items (authors).

1/65–73 **Machine Compilation and Editing of Printed Alphabetical Subject Indexes.** April 1964. Vladimir Slamecka. AMDO 15(2): 132–135.

1/65–74 **Boeing SLIP: Computer Produced and Maintained Printed Book Catalogs.** July 1964. E. A. Weinstein and J. Spry. AMDO 15(3): 185–190.

1/65–75 **The General Inquirer: A Computer System for Content Analysis and Retrieval Based on the Sentence as a Unit of Information.** Oct. 1962. Philip J. Stone et al. *Behavioral Science* 7(4): 1–15.

1/65–76 **Problems in Automatic Abstracting.** April 1964. H. P. Edmundson. CACM Vol. 7, No. 4, pp. 259–263.

A variety of problems concerning the design and operation of an automatic abstracting system is discussed. The purpose is to present a general view of several major problem areas. No attempt is made to discuss details or to indicate preferences among alternative solutions (author).

1/65–78 **A Technique for Computer Detection and Correction of Spelling Errors.** March 1964. Fred. J. Damerau. JACM 7(3): 171–176.

The method described assumes that a word which cannot be found in a dictionary has at most one error, which might be a wrong, missing or extra letter or a single transposition. The unidentified input word is compared to the dictionary again, testing each time to see if the words match—assuming one of these errors occurred. During a test run on garbled text, correct identifications were made for over 95 percent of these error types (author).

1/65–79 **Automatic Abstracting.** 1963. S. Szalai. *Idoszeru Muszaki Documentacios Kerdesek* (Current Problems of Technical Documentation), No. 5 In Hungarian.

The methods of probability and information theory applicable to automatic abstracting are discussed. The Luhn method of automatic abstracting is evaluated. It is shown that a great disadvantage of the latter is the need of a large computer memory to store the words which are of no descriptive value. The author proposes a method of automatically addressing and weighing relevantly frequent words which does not require a frequency check with preselected words of everyday language. The process is based on the linguistic properties of the Hungarian language which allow many trivial words to be identified by their length or grammatical construction. Investigations indicate that a similar method can be applied to flexible languages such as the Indo-German, and with certain modifications, to non-flexible languages such as English. The proposed method does not require, at least in its application to the Hungarian language, the truncation of words or complex grammatical transformations. The method is still experimental and needs considerable refinement. The author gives some sample abstracts prepared by this method (Doc. Inc.).

1/65–80 **Automatic Document Classification. Part II. Additional Experiments.** April 1964. Harold Borko and Myrna Bernick. JACM 11(2): 138–151.

This study reports the results of a series of experiments in the techniques of automatic document classification. Two different classification schedules are compared along with two methods of automatically classifying documents into categories. It is concluded that, while there is no significant difference in the predictive efficiency between the Bayesian and the Factor Score methods, automatic document classification is enhanced by the use of a factor-analytically derived classification schedule. Approximately 55 percent of the documents were automatically and correctly classified (authors).

Subject and Linguistic Analysis – General

1/65–81 **Rank Order Patterns of Common Words as Discriminators of Subject Content in Scientific and Technical Prose.** April 1964. Everett M. Wallace. System Development Corp., Santa Monica, California. 15 pp. SP–1505. AD–440 043.

There is a style of language characteristic of different subject areas which is particularly noticeable in scientific and technical writing. It is not only the unique vocabulary of a subject field which sets it apart from others, *but also the different habits of writers in using the most common words.* An experiment was devised to test whether these differences could be used for subject discrimination in addition to identification of unique vocabulary, particularly to determine whether or not author variation in style is sufficiently great to override the variation from field to field.

Fifty IRE abstracts in the field of electronic computers and fifty Psychological Abstracts were matched, one abstract at a time, one word type at a time, against two lists of words ranked in descending order of frequency as they occurred within two different sets of three hundred psychological and computer abstracts. All fully inflected forms of all function and content words were included in the rankings. Using the first 50 ranks only of the two lists, 93% of the abstracts were successfully discriminated. For the first 75 and 100 ranks, the success rates were 96% and 97%, respectively (author).

1/65–82 **Automatic Message Retrieval. Studies for the Design of an English Command and Control Language System.** Nov. 1963. Arthur D. Little, Inc., Cambridge, Mass. 187 pp. ESD–TDR–63–673. Final Report.

An associative searching technique has been developed for the automatic storage and retrieval of natural language message information. A user may employ his own vocabulary in formulating a typewritten request and the technique allows the machine to find relevant information even when there is no direct matching of vocabulary items. The method thus lends itself to ready interaction between computer and user. Answers are based on the network of associations implicit in the stored message data, and items are outputted in order of decreasing relevance to the question.

The technique can be advantageously implemented on existing data processing equipment and may eventually be used in conjunction with a novel form of associative memory-processor. The associative retrieval method does not require pre-indexing of the stored information. Also, its performance appears to improve as the amount of information stored increases and to be relatively insensitive to the nature of the message units employed. Finally, it should lend itself to eventual improvement when used in conjunction with language structure-utilizing techniques (author).

1/65–83 **Investigation of Systems for the Intellectual Organization of Information.** June 1964. Susan Artandi. Graduate School of Library Service, Rutgers, New Brunswick, N.J. 40 pp. USF grant GN–99.

The objective of the investigation is to bring together data on systems for the intellectual organization of information. Systems included in the study are described in terms of their dis-

tinctive characteristics, advantages, disadvantages, weaknesses, and limitations. It is assumed that a set of descriptive criteria applicable to information organization systems in general can be created and that sufficiently uniform descriptions can serve as a basis for comparative studies (author).

1/65–84 **Mechanized Indexing Studies of MSD Toxicity. Part II.** April 1964. John O'Connor. Institute for Scientific Information, Philadelphia, Pa.

A general procedure is described for using the meaning of a paper as a guide in looking for mechanized indexing rules. The procedure involves two basic concepts: 1. a paper saying something about T (toxicity, or any other index term being studied); and 2. an expression in a paper being evidence that the paper says something about T. Neither of these notions is completely well-defined, but either may be sufficiently intersubjective to help find useful mechanized indexing rules.

The application of the procedure to the *toxicity* papers in the sample described in Part I leads to the thesaurus keywords which were obtained more intuitively in Part I. It also leads to nine additional keywords, which function almost as well for mechanized indexing purposes. The general procedure leads to many other expressions which are also "toxicity clues," but which cannot be used directly for mechanized indexing because they are unlikely to recur in other *toxicity* papers. Study of these expressions suggests their generalization to "expression forms" containing variables. The possible values of the variables are defined for computer use by lists of "substance-contact words" and "disorder words." The feasibility of compiling useful forms of such lists is discussed.

Expression forms permit assigning *toxicity* mechanically to four *toxicity* papers (discussed in Part I) which contain no keywords. On the basis of expressions in *toxicity* and non-*toxicity* papers of the sample, various plausible-looking indexing rules using expression forms are suggested. The best of these, combined with rules involving keywords, selects all twenty-one *toxicity* (and "preferably *toxicity*") papers in the sample, at the cost of selecting nine other papers as well (author).

1/65–85 **Automatic Abstracting Evaluation Support.** Feb. 1964. Dan Payne and John F. Hale. Amer. Inst. for Research, Pittsburgh, Pa. RADC–TDR–64–30.

Techniques for producing two types of abstracts of technical literature were developed. One was designed to be a general-purpose abstract which could be used to support a variety of text-mediated tasks. The second was a modification of the first, and was designed to contain only that information necessary to the performance of a single given task. Four types of text-supported tasks were identified—screening, comprehension-retention, fact retrieval, and problem solving. Paper and pencil tests were prepared—the tests designed to serve as analogs of the four tasks. Performance accuracy and performance time on the tasks of subjects using the abstracts were compared to those of subjects using original text. For both college students and professional scientists, it was found that:

1. The abstracts effectively served in place of full text in supporting performance on all tasks except fact retrieval. Accuracy loss was slight, while time savings were substantial. The abstracts were more effectively used for screening and comprehension-retention than for problem solving.
2. The task-specific abstracts were superior to the general-purpose abstracts in amount of time saved in performance of the screening and the comprehension-retention tasks.

The implications of these results for future work were discussed (authors).

1/65–86 **Relational Indexing and New Methods of Concept Organization for Information Retrieval.** 1963. J. Farradane. *Automation and Scientific Communication,* pp. 135–136 (see 7/64–151).

Relational indexing, which has proved valuable in detailed indexing for retrieval, requires complementation by new methods of concept organization and control, which are now under investigation. Research on the nature of the variables involved and on tests methods and their statistical validity is now being planned (author).

1/65–87 **A New Efficient Structure-Matching Procedure and Its Application to Automatic Retrieval Systems.** 1963. Gerard Salton, E. H. Sussenguth, Jr. *Automation and Scientific Communication,* pp. 143–146, (see 7/64–151).

A new automatic method is presented for the comparison of two-dimensional line patterns. Retrieval applications include the matching of chemical structures, the comparison of syntactically analyzed excerpts extracted from documents and search requests, and the matching of document identifications consisting of two-dimensional graphs with query identifications (author).

1/65–88 **English-Like Systems of Mathematical Logic for Content Retrieval.** 1963. Herbert G. Bohnert. *Automation and Scientific Communication,* pp. 155–156 (see 7/64–151).

An English-like system of mathematical logic is a formally defined set of sentences whose vocabulary and grammar resemble English, with an algorithm which translates any sentence of the set into a notation for mathematical logic. Objectives, accomplishments, and problems in the construction of such languages in Project LOGOS are discussed (author).

1/65–89 **Some Search Strategy Factors in Automated Searches of Technical Reports at DDC.** 1963. J. Heston Heald. *Automation and Scientific Communication,* pp. 159–160 (see 7/64–151).

The relationship between subject-matter controls and search methods is discussed. The DDC philosophy is explained for treating descriptors with two levels of importance and for their built-in roles and links, through thesaurus usage and coding techniques (author).

1/65–90 **Conceptual Analysis of Questions in Information Retrieval Systems.** 1963. Allan M. Rees. *Automation and Scientific Communication,* pp. 175–178 (see 7/64–151).

The nature of the question-asking process is described in terms of a series of evolutionary steps from information problem —information need—formulation of question—analysis by searcher—negotiation—definition—enumeration of search concepts —translation of concepts into indexing language—selection of search strategy. The process is believed to be true of questions put to all retrieval systems (author).

1/65–91 **The Role of Thesauri in the Convergence of Word and Concept Indexing.** 1963. Saul Herner. *Automation and Scientific Communication,* pp. 183–184 (see 7/64–151).

The increase in indexing vocabularies guiding the use of terms and indicating their relationship has resulted in a convergence of word and concept indexing. Vocabularies have emerged as devices for converting from words to concepts. The intellectual role of the indexer has shifted to the vocabulary designer. Current manifestations of this shift are discussed (author).

1/65–92 **The Relation of Requests to Concepts of File Organization and Search.** 1963. Ralph M. Shoffner. *Automation and Scientific Communication,* pp. 187–188 (see 7/64–151).

A characterization of the relation of requests to file organization and search is presented. The requirement for exact-match or near-match response and the relation of processing effort to request predictability determine file organization. Desired results and allowable processing effort determine the search rules (author).

1/65–93 **Thesaurus for Literature Analysis in a Cancer Research Institution: Construction and Correlation with Three National Thesauri.** 1963. Lois F. Lunin, Marily Stovall. *Automation and Scientific Communication,* pp. 189–190 (see 7/64–151).

In constructing a theasaurus for computer-oriented literature retrieval activities, two departments of M.D. Anderson Hospital found that: 1. cooperative interinstitutional effort was practical (48 percent of terms were duplicates); 2. although compatibility was maintained as far as possible, national thesauri were inadequate for researchers who use and generate information (authors).

1/65–94 **Paragraph Analysis Novel Technique for Retrieval of Portions of Documents.** 1963. S. N. Jacobson. *Automation and Scientific Communication,* pp. 191–192 (see 7/64–151).

A significant similarity between the retrieval of entire documents and the retrieval of selected portions of documents is described. Based on this similarity, procedures for segmenting documents into independently retrievable segments and assigning index terms to these segments are presented (author).

1/65–95 **Research on a Linear Network Model and Analog Device for Associative Retrieval.** 1963. Paul E. Jones. *Automation and Scientific Communication,* pp. 211–212 (see 7/64–151).

In investigations at Arthur D. Little, Inc., the applicability of linear network analysis to the retrieval of information from a given body of data has been pursued. This point of view provides a framework which encompasses many of the parameters of the retrieval process and enables analytic exploration of the model (author).

1/65–96 **Analysis, Indexing and Correlation of Information.** 1963. K. H. Meyer-Uhlenried, G. Lustig. *Automation and Scientific Communication,* p. 229 (see 7/64–151).

Research within EURATOM-CETIS refers to comparison of text with thesauri, keywork-in-context method, and automatic classification on the basis of statistical evaluation by comparison of text with hierarchically organized keyword lists (author).

1/65–97 **A Facet Analysis System.** 1963. R. S. Hooper. *Automation and Scientific Communication,* pp. 253–254 (see 7/65–151).

1/65–98 **Mathematical Theories of Relevance with Respect to Systems of Automatic and Manual Indexing.** 1963. Donald J. Hillman. *Automation and Scientific Communication,* pp. 323–324 (see 7/64–151).

Several mathematical theories are described to overcome the serious practical problems of identifying the relatedness of documents in a collection, constructing metrics to define the degree of relatedness and measuring relevance of documents to queries. A general theory of indexing is based on models which are experimentally corroborated (author).

1/65–99 **Automatic Document Classification Using Information Theoretical Methods.** 1963. Alfred Trachtenberg. *Automation and Scientific Communication,* pp. 349–351 (see 7/64–151).

This paper presents a technique, using two information theoretical measures, of selecting clue words whose occurrence in a document would then be the basis for the automatic classification of that document. Experiments are proposed on data used by Maron and by Borko and Bernick, to compare the effectiveness of the three methods (author).

1/65–100 **A Technique for Determining Index Requirements.** April 1964. C. Jahoda. AMDO 15(2): 82–85.

1/65–101 **The Consistency of Human Judgments of Relevance.** April 1964. A. Resnick and T. R. Savage. AMDO 15(2): 93–95.

1/65–102 **Correlation of Indexing Headings and Title Words in Three Medical Indexing Systems.** J. O'Connor. April 1964. AMDO 15(2): 96–104.

1/65–103 **The Distribution of Term Usage in Manipulative Indexes.** April 1964. Nona Houston and Eugene Wall. AMDO 15(2): 105–114.

1/65–104 **Detailed Discussion of Bar-Hillel's "Theoretical Aspects of the Mechanization of Literature Searching."** April 1964. Ron Manly. AMDO 15(2): 126–131.

1/65–105 **Chemical Documents and Their Titles: Human Concept Indexing vs. KWIC Machine Indexing.** April 1964. Mary Jane Ruhl. AMDO 15(2): 136–141.

1/65–106 **Identification of Research Papers by the Physical Units Appearing in the Paper.** April 1964. Pauline Atherton. AMDO 15(2): 150.

1/65–107 **Indexing and Dependency Logic for Answering English Questions.** July 1964. R. F. Simmons, S. Klein, and K. McConlogue. AMDO 15(3): 196–204.

1/65–108 **Correlative Indexes. IX. Vocabulary Control.** April 1964. Charles L. Bernier. JCHD 4(2): 99–103.

Control of subject-indexing vocabulary brings together like information and aids word location. Twelve means of control are discussed: subject area, cross references, organization of terms, definition and differentiation of terms, number of indexing terms in vocabulary, hospitality for new terms, specificity of indexing terms, semantic versus statistical use and syntax of terms, compatibility with other vocabulary, choice of terms used to represent subject indexed, subject importance of index terms chosen, and roles and links. Examples illustrate means (V.E.Y.).

1/65–109 **Correlative Indexes. X. Subject-Index Qualities.** April 1964. Charles L. Bernier. JCHD 4(2): 104–107.

Principal subject-index qualities are completeness and consistent organization. Others are: guidance, subject indexing, format, typography, and price. Incomplete indexing — discounting money limitation or lack of subject knowledge by indexer — can be corrected by indexer-training and verification of indexing. Carefully edited thesaurus prevents scattering of like information. Format — unless very poor — has negligible effect on usefulness of an index; typography affects legibility. Cost of indexes should be evaluated against cost of duplication of research and loss of scientists' time (V.E.Y.).

1/65–110 **Fact Retrieval and Deductive Question-Answering Information Retrieval Systems.** April 1964. William S. Cooper. JACM 11(2): 117–137.

Information retrieval systems may be classified either as Document Retrieval systems or Fact Retrieval systems. It is contended that at least some of the latter will require the capability for performing logical deductions among natural language sentences. The problem of developing systems of logical inference for natural languages is discussed, and an example of such an analysis for a sublanguage of English is presented. An experimental Fact Retrieval system which incorporates this analysis has been programmed for the IBM 7090 computer, and its main algorithms are stated (author).

1/65–111 **Keywords and Clumps.** March 1964. R. M. Needham and K. Sparck Jones. JDOC 20(1): 5–15.

Recent work on information retrieval at the Cambridge Language Research Unit aimed to develop automatic subject indexing procedures. Because of difficulties in setting up a system

of keywords taken from the text of documents, a controlled structure of these words used for coordinate indexing was developed for 800 offprints on the subject of automatic language processing. Persons chose terms, arranged them in a multiple hierarchy and assigned numbers to each term; optical-incidence cards with one card per term revealed hits and clusters of terms which have "a strong mutual tendency to co-occur" were determined, and the system was then mechanized on an EDSAC-2 computer to reduce the labor in application. With co-occurrence coefficients for all the pairs of terms, classes of mutually co-occurring terms could be found using the definition: "An acceptable class (or *clump*) is a subset of the objects such that each member of the subset has a greater aggregate of coefficients to the other members than to the non-members, the reverse holding for all non-members." An example of one class result is: phrase marker, kernels, string operations, transformational grammar, terminal language, Markov process, finite state process. In indexing, one may then assign a document to several classes, as if using several descriptors (An abstract thesaurus could be used to create "proto-clumps" with less weight than document derivations.) In retrieving, keywords as well as classes are used since keywords alone do not give as good results. It is indicated that fuller discussion of the theoretical and programming aspects of this NSF-supported work is given in other CLRU publications (D.C.W.).

1/65–112 **Communication of Science Information.** May 10, 1963. Percy H. Tannenbaum. SCIE. Vol. 140, No. 3567, pp. 579–583.

1/65–113 **The Reference Function of the Catalog: Some Questions Concerning Responsibility.** Spring 1964. Robert M. Pierson. LRTS 8(2): 153–156.

Makes various points for and against the creators and users of catalogs in their concern with its form. One point suggests that conflict is caused by a tendency in library education to train in the "construction" of catalogs and in the "interpretation" of other reference tools on the mistaken assumption that one who has learned to catalog will automatically become a competent interpreter of catalogs (D.C.W.).

1/65–114 **Indexing Theory, Indexing Methods, and Search Devices.** 1964. Frederick Jonker. Scarecrow Press, Inc., New York. 124 pp. $4.00.

Specific Classification and Indexing Systems

1/65–115 **The Classified Catalog at Boston University, 1948–1964.** Summer 1964. Mary Darrah Herrick. LRTS 8(3): 289–299.

This catalog, using the Library of Congress classification, is believed to be unique. Discussion covers the reasons for its selection, how it was started, difficulties that have arisen, benefits that have appeared, the preparation and evaluation of its Index, cost estimates in terms of staff and volumes processed, the use of the catalog at BU, and the reasons for satisfaction with the LC classification as the catalog base. It is surmised that, when library catalogs are computer generated, the classification numbers will be easier to feed into the machine than will subject headings (D.C.W.).

1/65–116 **Toward the Seventeenth . . . Dewey.** Spring 1964. Sarah K. Vann. LRTS 8(2): 172–187.

This report summarizes the "vulnerability and durability" of the Dewey Decimal Classification, describes the use of the 16th Dewey by thirty libraries in Texas, and indicates particular points where changes can be anticipated in the 17th. Concerning relocations, it is reported that a major deterrent to their acceptance is the feeling of responsibility to reclassify for consistency, especially because of the open shelf concept. Book retirement programs and the increased size of libraries are said to vitiate the argument; and librarians know that only through well-designed catalogs can the total library resources be made available. The author quotes Kenneth Povey from the *Journal of Documentation*, 1946, that browsing may be regarded objectively as an "uncivil waste of readers' time and a repudiation of exact scholarship." Among anticipated changes are expansion of some two dozen subjects, general re-use of numbers only if they have been vacated for twenty-five years, and creation of a new "Table of Standard Subdivisions" and an area table (D.C.W.).

1/65–117 **A Novel Index Tailored to Plastics Specialists.** July 1964. A. M. Anzalone, C. A. Brokars, and G. Cohn. AMDO 19(3): 191–195.

1/65–118 **U.D.C. Abridged Russian Edition.** July 1964. A. Kutten. AMDO 15(3): 226.

Coding and Notation

1/65–119 **National Academy of Sciences-National Research Council. Survey of Chemical Notations.** A report of the Committee on Modern Methods of Handling Chemical Information, Division of Chemistry and Chemical Technology, NAS-NRC. (NAS-NRC Pub. no. 1150), Washington, NAS-NRC, 1964. 495 pp. AD-600 920. Available from Printing and Publishing Office, NAS-NRC, 2101 Constitution Ave., Washington, D. C. $7.00.

In recent years, the problem of identifying complex chemical compounds by conventional chemical means has become so difficult that various codes and ciphers have been devised for the purpose. The ones that have found the most use—known as linear notation systems—consist of single-line sequences of letters, numbers and other common typographical symbols. When perfected, such systems can identify unambiguously each one of some three million known chemical compounds. Besides the better known systems—such as the Wisswesser and IUPAC notations, which are aimed at complete applicability—many simpler systems have been devised for limited use.

The purpose of the above report is to survey all the chemical notation systems known to be operating or under development in the United States during the period May, 1961, to July, 1962 and to make recommendations for the future concerning the application of modern methods to the rapidly growing problems of processing and using chemical information. Included in the survey are reports on fifty site visits, together with summaries and analysis of the findings, and an extensive bibliography of 328 items. A supplement to the report covering notational systems in use in Europe and Japan is expected to be issued in the near future. (J.A.W.).

1/65–120 **Encoding of Organic-Chemical Structural Formulas and Reactions by Machine.** 1963. Ernst Meyer. *Automation and Scientific Communication*, pp. 131–132 (see 7/64–151).

A method and a device are described by means of which it is possible to bring structural formulas and reactions into a form suitable for machine processing, without loss of information, so that they can be handled by Automatic Data Processing Systems and automatically transformed into other coding systems (author).

1/65–121 **Use of Meta-Language in Information Retrieval Systems.** 1963. W. Goffman, J. Verhoeff, Jack Belzer. *Automation and Scientific Communication*, pp. 181–182. (see 7/64–151). Also available as AFOSR-J1538 and AD-427 958.

1/65–122 **Automatic Coding of Chemicals Directly from Structure Pictures.** 1963. Robert S. Ledley. *Automation and Scientific Communication*, pp. 201–202 (see 7/64–151).

1/65–123 **Cross-Reference Index Identifying Range of Scientific Journal Titles Represented by Abbreviated Citations.** 1963. Phyllis A. Richmond. *Automation and Scientific Communication*, pp. 225–226 (see 7/64–151).

Journal title citations are so abbreviated and so ambiguous that identification of original source journals is difficult. An index is being compiled to provide *all* the full titles which may be pertinent to a given citation. A citation formula is sought which will identify each journal as a unique entity (author).

1/65–124 **The Naming of Journals and Organization: Implications for Document Recovery.** 1963. Robert L. Birch. *Automation and Scientific Communication*, pp. 307–308 (see 7/64–151).

Names chosen for technical journals for societies often incorporate factors which needlessly complicate filing and recovery. Changes of name also appear to often ignore the effect on commonplace information retrieval procedures. Factors considered include ease of memory retention, compatibility of wording and of layout of title pages with filing systems used in offices, libraries, and bibliographies (author).

1/65–125 **Bibliographic Encoding: A Machine-Interpretable Natural Format for Highly Structured Data.** 1963. Paul J. Fasana. *Automation and Scientific Communication*, pp. 325–326. (see 7/64–151).

A method for encoding bibliographic data for machine interpretability is outlined. The encoding method uses a "machine-interpretable" natural format (MINF) which is defined as a variable length field, multilevel, sequential format in which natural typing operations are used as machine-identifiable boundary codes. A practical example of MINF adapted to a conventional library catalog card is presented (author).

1/65–126 **The Application of Random Superimposed Coding and Chain Spelling to Peek-A-Boo Cards.** April 1964. W. Uhlmann. AMDO 15(2): 89–92.

1/65–127 **A Proposed Improvement in the Printing of Chemical Structures, Which Results in Their Complete Computer Codes.** July 1964. A Feldman. AMDO 15(3): 205–209.

1/65–128 **Punch Card File of Characteristic Values of Chemical Compounds.** 1964 E. Offermann, and A. Burkhardt. DKMN No. 2, pp. 34–38. In German.

An overlapping punch card system with Universal Decimal Classification code is described for filing data on characteristic values of chemical compounds. The UDC numbers of chemical compounds are punched using a triangular code based on a punching field with four pairs of holes. Terms describing chemical and physical properties of particular compounds are coded directly. Other data is punched using a double pair code with four punching positions. The system is being operated with good success by an East German information service (Doc. Inc.).

1/65–129 **Coding of Punch-Cards, Especially Edge-Notched Cards.** H. Schuetz. DKMN No. 2, pp. 43–47. In German.

Coding in binary and ternary numerical systems is described. The codes provide for compact information storage on edge-notched cards. Holes are punched by notching and slotting or by notching alone. The methods described are not suitable for superimposed coding (author).

1/65–130 **Generating a Canonical Prefix Encoding.** March 1964. Eugene S. Schwartz and Bruce K. Allick. JACM 7(3): 166–169.

Computer programs for generating a minimum-redundancy exhaustive prefix encoding are described. One program generates a Huffman frequency tree, another determines the structure functions of an encoding, and a third program assigns codes (authors).

1/65–131 **Symposium on Inorganic Nomenclature. Introductory Remarks.** 1964. Janet D. Scott. JCHD 4(2): 63–64.

1/65–132 **Procedures in the Development of Chemical Nomenclature.** 1964. E. J. Crane. JCHD 4(2): 64–66.

1/65–133 **Inorganic-Organic Nomenclature.** 1964. Janet D. Scott and Kurt L. Loening. JCHD 4(2): 66–69.

This paper points out some of the similarities and some of the differences between organic and inorganic nomenclatures and examines problems briefly in order to contribute in a small way to a greater mutual understanding between chemists active in inorganic as well as organic chemistry (P.M.B.).

1/65–134 **Nomenclature of Coordination Compounds: Present Status.** 1964. W. Conard Fernelius. JCHD 4(2): 70–84.

Even the most casual glance at the current journals indicates the rapid rate at which the chemistry of the coordination compounds is developing and expanding. It is to be expected that this evolving field of chemistry will continue to bring new nomenclature problems. The study of nomenclature must keep pace with the new developments. For that reason, it is essential that broad patterns capable of extension be followed so that the new problems can be solved without breaking so completely with previous practices as to render the earlier literature not readily understandable (55 references) (P.M.B.).

1/65–135 **Nomenclature of Phosphorus Compounds.** 1964. John R. Van Wazer. JCHD 4(2): 84–90.

Probably a greater number and variety of compounds have been described for phosphorus than for any other element, aside from carbon. Furthermore, the study of phosphorus compounds is a very active branch of present-day chemistry, and knowledge of phosphorus chemistry is widening. Although the chemistry of phosphorus has been reasonably well systematised, the present-day nomenclature is most inadequate and often confusing. Indeed, when one is writing about any organic phosphorus compound, except a few of the more common ones, it is wise to have a structural formula accompany the name when it is first stated. Furthermore, phosphorus chemistry during the past decade or so has been extended to include whole families of compounds in which phosphorus atoms appear repeatedly in the molecular backbone, and there must be names not only for individual compounds but also for families. Specific problems and solutions are indicated (P.M.B.).

1/65–136 **Some Problems in Halogen Nomenclature.** 1964. A. F. Clifford. JCHD 4(2): 91–95.

A progress report. The "halogen" system (described) or more especially the "phosphorus" system of nomenclature is much more adaptable than "classical" inorganic nomenclature to most of the naming problems likely to arise for the more complex compounds of the halogens. It also provides systematic names for the simple compounds from which the complex ones must inevitably be derived. It is recognized, of course, that these names would undoubtedly not displace the common (trivial?) ones for simple compounds (e.g., chloruanoic acid for chlorous acid) any more than ethanoic acid has replaced acetic acid. On the other hand, the parent name, chloruane, provides a basis for a truly systematic name for its more complex derivatives, whereas the name chlorous acid does not. Furthermore, such nomenclature will be compatible in a complex name with the nomenclature of the organic groups with which it must frequently be associated, not only as side chains, but as backbone constituents. This also cannot be said for classical inorganic nomenclature (P.M.B.).

1/65–137 **Nomenclature Problems in Boron Chemistry.** 1964. Roy M. Adams. JCHD 4(2): 95–98.

A number of current problems in boron nomenclature are listed. Nomenclature groups on both the national and international levels are seeking the best solutions. As chemistry grows, so must nomenclature standards to prevent chaos in communication. Several examples are given (P.M.B.).

1/65–138 **Peek-A-Boo Card System for Author File, JK.** 1964. *Metodika a Technika Informaci* (Methodology and Technique of Information), No. 2, pp. 43–45. In Czech.

Disadvantages of existing systems are explained. It is proposed that peek-a-boo cards be used to encode the names of authors. The cards would have 1,026 punching positions that are identifiable by direct printing of all possible combinations of the first three letters of any author's name. The letters are so selected as to give a well balanced distribution of the entries, too. A set of such letter combinations to identify punching positions is included (Doc. Inc.).

Equipment and Material Descriptions

1/65–139 **Microfiche, a New Information Media.** 1963. C. P. Yerkes. *Automation and Scientific Communication*, p. 129 (see 7/64–151).

1/65–140 **The Design of an Information Storage and Retrieval System: FMA's Filesearch.** 1963. R. A. Condon. *Automation and Scientific Communication*, pp. 137–138 (see 7/65–151).

1/65–141 **An Electronic Retina Optical Character Reading System with Integral Editing and Indexing Capabilities.** 1963. Thomas Q. LeBrun. *Automation and Scientific Communication*, pp. 149–150 (see 7/64–151).

1/65–142 **A Selected Group of Techniques for Automated Microimage Handling.** 1963. Brien O'Brian, Jr. *Automation and Scientific Communication*, pp. 197–198 (see 7/64–151).

1/65–143 **The Rapid Selector as Currently Used for Information Search and Replica Copy Retrieval.** 1963. Thomas C. Bagg. *Automation and Scientific Communication*, pp. 227–228 (see 7/64–151).

There are two Rapid Selector installations. One is at the National Bureau of Standards being used to study effectiveness of indexing and coding methods for machine retrieval. The other is at BuShips being used for handling of specification, instruction manuals, engineering reports, etc. (author).

1/65—144 **A Read-Only Multi-Megabit Parallel Search Associative Memory.** 1963. G. G. Pick, D. B. Brick. *Automation and Scientific Communication*, pp. 245–246 (see 7/64–151).

A fully parallel associative ready-only memory capable of storing megabits and operating at over a hundred thousand interrogations per second is described. The solenoid array memory stores the data on mechanically punched data sheets through which the solenoids pass. It is particularly applicable to word translators and indexing structures (author).

1/65–145 **Adaptive and Learning Recognition Machine to Produce Codes, or Punch Cards in Response to Spoken Words.** 1963. Arnold Lesti. *Automation and Scientific Communication*, pp. 279–280.

CYNTHIA III is a scientific demonstrator of machine learning techniques and systems as applied to recognition of live speech, regardless of speaker, and other forms of highly statistical signals. Besides demonstrating the recognition of any phoneme, accents, and speaker, the machine has recognized seismic signals which cannot be separated by any other known method (author).

1/65–146 **Electronic Retina Character Reader for Data Input.** 1963. Herman L. Philipson, Jr. *Automation and Scientific Communication*, pp. 285–286 (see 7/64–151).

1/65–147 **New Photon Composing Machine Will Set Tape from the Output of a Computer.** 1963. R. G. Crockett. *Automation and Scientific Communication*, pp. 317–318 (see 7/64–151).

"ZIP" is a printout device that will produce justified text of typographical quality on film or paper from the off-line operation of a computer. Character and command codes are read at high speed from magnetic tape; processed in a control console, and then are used to initiate the projection of characters onto film (author).

1/65–148 **The Coordinate Concept: An Approach to Tape Punching.** August 1964. D. B. Holland et al. *Control Engineering*, pp. 60–62.

The unique position that each character on a typewriter page occupies can be defined by X–Y coordinates: space 26, line 10, for example. Basing their design on this concept, the authors have developed a prototype tape typewriter on which punched tape can be prepared fast with ease. The restriction that conventional typewriters put on the typist—she is not allowed to move back to previous lines on the page—is removed, thus giving her virtually two degrees of freedom in the vertical direction (authors).

1/65–149 **Microfilm Readers and Reader-Printers.** Feb. 1964. Don Boyett. NAMN, No. 68, pp. 174–188 (A review and comparative specifications for 43 models).

1/65–150 **Activate your Microfiles with Unitized Microforms.** April 1964. Frank L. Hilton, Jr. RPRV 14(4): 56–62 (includes description of Recordak microfiche and new reading equipment).

1/65–151 **New Wrinkle on Retrieval from the Microfilm Reel.** April 1964. Phil Hirsch. RPRV 14(4): 19–23 (description of the operation of the Bureau of Ships Rapid Selector).

Use and User Studies

1/65–152 **A Use Evaluation of the AIChE Bibliographic Descriptions and the Chemical Engineering Thesaurus.** May 1, 1964. Saul Herner, W. F. Johanningsmeier, and D. T. H. Campbell. A report to the National Science Foundation, Washington, D. C., Herner and Company.

A mailed questionnaire survey and follow-up telephone interviews were employed to obtain detailed information on mode, purpose and context of use, and reasons for non-use, of the AIChE *Thesaurus* and bibliographic descriptions. Slightly over a third of the respondents to the questionnaire (648 out of 1,915) had made some use of the bibliographic descriptions, four-fifths of these for current awareness rather than for retrospective searching. Only 139 of the 1,915 respondents had made use of the *Thesaurus*, the principal application being a guide to indexing and filing. It had also been used for comparison with other vocabularies and in the study of indexing authorities. It was infrequently employed as a guide in the performance of searches. Lack of interest or need was the main reason given for non-use of both tools. In general, the non-users of these tools were also light users of the chemical engineering literature and of indexing and searching aids. Only 15% of the respondents were familiar with links or roles, and only 25% with concept coordination indexes. Three-quarters of the non-users of the *Thesaurus* had never seen a copy. The general conclusion reached by the investigators was that lack of use is a reflection less on the tools themselves than on the way they have been presented and publicized (authors).

1/65–153 **Bibliography of Use Studies.** March 1964. Richard A. Davis and Catherine A. Bailey. A report by the Drexel Institute of Technology, Graduate School of Library Science, Philadelphia, Pa. 98 pp. 438 refs. AD–435 062. Grant GN170, Proj. 195.

1/65–154 **Reports of the American Psychological Association's Project on Scientific Information Exchange in Psychology—Volume 1.** Dec. 1963. American Psychological Assoc. Wash., D. C. 283:. Reports of work performed under NSF grant G–184/64.

This volume contains the reports produced during 1963 by the APA Project on Scientific Information Exchange in Psychology. Seven of these reports, APA–PSIEP Reports No. 1 through No. 7, were distributed in technical report form in August, 1963. The first report in this volume, APA–PSIEP Overview Report B, which describes the rationale of the work and summarizes the results, is a revision of APA–PSIEP Overview Report A distributed in August, 1963. Two reports (No. 8, 9) contained in this volume have not been distributed before.

CONTENTS:

B. An Overview of the Structure, Objectives, and Findings of the American Psychological Association's Project on Scientific Information Exchange in Psychology. December 1963 (see 4164–11).

1. Scientific Activity and Information Problems of Selected Psychologists: A Preliminary Survey, August 1963 (see 4/64–66).
2. An Informal Study of the Preparation of Chapters for the *Annual Review of Psychology*. August 1963 (see 4/64–67).
3. A General Study of the Annual Convention of the American Psychological Association. August 1963 (see 4/64–68).
4. Convention Attendants and Their Use of the Convention as a Source of Scientific Information. August 1963 (see 4/64–69).
5. Convention Participants and the Dissemination of Information at Scientific Meetings. August 1963 (see 4/64–70).
6. Publication Fate of Formal Presentations at the 1957 Convention of the American Psychological Association. August 1963 (see 4/64–12).
7. Archival Journal Articles: Their Authors and the Processes Involved in Their Production. August 1963 (revised December 1963) (see 4/64–71).
8. A Comparison of Scientific Information-Exchange Activities at Three Levels of Psychological Meetings. December 1963.
9. The Use of Scientific Journals by Psychologists and the Readership of Current Journal Articles. December 1963.

1/65–155 **Searchers Want Facts Not Folios—Retrieve Data Not Documents—TheNeedle Is Dull—Sharpen It with Automation.** 1963. W. H. Waldo. *Automation and Scientific Communication*, pp. 207–209 (see 7/64–151).

Storage of laboratory data available through appropriate questions to a computer has been successfully tested by Monsanto as the optimum replacement for the uncommunicative classical communications system of report writing by hand, storage in file drawers and retrieval by key words. The preparation of review articles is theoretically more desirable, but apparently not practical (author).

1/65–156 **Density of Library Book Use in Biology and Medicine.** 1963. Frederick G. Kilgour. *Automation and Scientific Communication*, pp. 263–264 (see 7/64–151).

Use quotients were obtained by dividing numbers of book volumes published in 1951–62 and circulated in one year by volumes in the Library's corresponding subject classes. Books of the last dozen years experienced six times as concentrated use as the entire book and journal collection (author).

1/65–157 **Some Human Elements in the Handling of Technical Information—Reflections of an Information Scientist.** 1963. G. E. M. Wohlauer. *Automation and Scienfic Communication*, pp. 281–282 (see 7/64–151).

1/65–158 **A Survey on the Information Practices of Engineers at Western Electric.** April 1964. N. P. Levy. AMDO 15(2): 86–88.

1/65–159 **Research on Users' Needs: Where Is It Getting Us?** Feb. 1964. C. W. Hanson. ASLP 16(2): 64–78.

1/65–160 **Scientific Information: Problems and Prospects.** Autumn 1963. B. C. Vickery. *Minerva* 2(1): 21–38.

Scientists need information both in order to keep abreast of published and *not yet published* work, and to avoid repeating research already done. They also need specific items of information from time to time or exhaustive surveys of a special subject. Fast communication among scientists is usually based on informal and somewhat haphazard methods rather than on publication, although the latter is much to be preferred if it can be speeded up. There has been a proliferation of sciences, scientists and scientific publication; subjects, people and journals are widely scattered. So far nothing is as satisfactory as print for effective, unified communication. Review articles and journals and abstracting and indexing journals have grown up to fill deficiencies caused by scattering. Considerable duplication of effort occurs among them although they are addressed to different audiences. Waste of labor could be avoided by use of a single author abstract as suggested by I.C.S.U. Abstracting Board. Speedier publication of abstracts is needed. Indexing of abstract journals needs to be improved.

Application of computers and other machines to dissemination of scientific information promises future benefits, although present applications are limited. "Of the three basic types of documentation—collection, indexing and search—it is indexing that presents the most difficult and costliest procedure." Scientists still have to read items containing information. Libraries are indispensable in the processes of collecting, storing and servicing media of information. Cooperation among libraries is needed because no one can be self-sufficient. Current, significant serial literature should be widely held. Older scientific material can be stored in central depositories. Comprehensive coverage in specialized centers would allow local institutions to cater to local needs.

Concerning personnel, capability in science does not ensure capability in use of scientific literature. British practice is to provide technical liaison officers between industry and libraries or information officers within an organization. High caliber documentation personnel are badly needed for developing more adequate approaches to research and development in documentation, particularly in basic thinking. If these approaches are to succeed, more scientists will have to commit themselves seriously to the work of information handling (P.A.R.).

1/65–161 **Scientific-Technical Information and Its Utilization.** 1964. I. Spejzl. *Podnikova Organizace* (Factory Organization) No. 7, pp. 303–305. In Czech.

Factors that affect the volume and quality of scientific information and its use in the Czechoslovak machine industry are discussed. The fundamental factors affecting the volume and quality of scientific information are: 1. the quantity, experience, capabilities, and choice of scientific personnel; 2. the methodology and organization of scientific work, including the use of recorded scientific information; 3. the availability and quality of scientific equipment and instrumentation; and 4. the length of time devoted to research. If the overall volume and quality of information is expressed as a product of values (between zero and 1) assigned to each factor, the weakest factor exercises the most pronounced effect; hence, the optimum value is obtained if all factors are balanced among themselves. The utilization of information varies with the ratio of research on immediate problems to that on perspective problems; at the State Research Institute for Material and Technology in Prague, this ratio is 2:1 (the prewar ratio was 1:2). Use of scientific information is said to decrease linearly with its "age"; the author shows that as a result of the time lag between research conclusion and technical production, the actual "economic effectiveness" of information is not more than three years for information with theoretical life span of five years (Doc. Inc.).

1/65–162 **Investigation of Special Literature Requirements in a Planning Office.** 1964. F. Patek. *Tudomanyos es Muszaki Tajekoztatas* (Scientific and Technical Information), No. 4/5, pp. 275–292. In Hungarian.

Research on improvement of information services of special libraries for engineers and technicians in planning offices is reported. Questionnaires were distributed to 472 engineers and technicians. The problem was whether company and technical public libraries satisfy the need for technical information. The following experiences were reported: the services of the library would be more valuable if books would be lent for an unlimited period of time; open stack system is preferable; and more copies of books should be available. Many specialists stated that they had to set up private libraries at home and study after office hours because of inadequate services of company of public libraries. A surprising disclosure of the investigation was that most of the specialists very rarely utilized periodicals for information (Doc. Inc.).

1/65–163 **Specialists and Special Literature.** 1964. N. Hegyi. *Tudomanyos es Muszaki Tajekoztatas* (Scientific and Technical Information), No. 4/5, pp. 309–319. In Hungarian.

The use of technical-scientific literature by specialists and skilled workers is analyzed and measures are proposed to stimulate interest in such literature. Only 4.07 percent of workers use the services of technical libraries, and the percentage attending technical training courses is even smaller. It is recommended that more funds be allotted for libraries and enterprises to disseminate technical information, to publish more technical books, and to diversify the methods of technical training (Doc. Inc.).

Design, Testing, and Evaluation of Information Systems or Services

1/65–164 **Information Retrieval and Subject Indexing: Cranfield and After.** June 1964. Barbara R. F. Kyle. JDOC 20(2): 55–69.

The Cranfield project demonstrated that, *other things being equal,* the same operating efficiency can be achieved by the use of alphabetical indexing, uniterms, UDC or a faceted classification. Some techniques are being investigated to see how far efficiency may be increased. Several suggestions are put forth of areas for further research (D.C.W.).

1/65–165 **Time Standards for Operations in the National Library in Mostar.** 1964. Z. Hamzic, *Bibliotekarstvo* (Library Science), No. 1, pp. 51–58. In Serbo-Croatian.

This is a detailed list of elementary library operations and the amount of time required for these operations. The list begins with the registration of a reader (estimated at four minutes), checking in books (two minutes), checking out books including short conversation with the reader (four minutes), replacing the books on the shelf (forty seconds), etc. More than one hundred operations are described and the time is averaged. These data are interpolated to the number of readers per day or per month (Doc. Inc.).

1/65–166 **Linear vs. Inverted File Searching on Serial Access Machines.** 1963. R. E. Swid. *Automation and Scientific Communication*, pp. 169–170 (see 7/64–151).

Based on a comparison of: 1. existing systems; 2. tape processing; 3. internal functions; and 4. file maintenance problems, a linear file organization is found to be superior to an inverted one for searching on tape oriented computer systems. Programming sophistication and machine configuration can change this relationship only in degree (author).

1/65–167 **Is Relevance an Adequate Criterion in Retrieval System Evaluation?** 1963. Lauren B. Doyle. *Automation and Scientific Communication*, pp. 199–200 (see 7/64–151). SP-1262. AD-434 484.

It is argued that the use of "relevance to a search request" as a criterion of what a system retrieves is, in effect, a suboptimization on the machine side of the man-machine interface, and that the searcher needs an efficient exploratory system rather than a request-implementing system (author).

1/65–168 **Feasibility Criteria for Establishing Specialized Information Centers.** 1963. John Sherrod. *Automation and Scientific Communication*, p. 221 (see 7/64–151).

Some guidelines are presented for estimating the practicability of establishing specialized information centers. To arrive at a valid justification for a new service, criteria will need to be developed in each of four areas: clientele to be served; information to be handled; provision of adequate funding; and, relationships to existing services (author).

1/65–169 **Design of an Experiment for Evaluation of the Citation Index as a Reference Aid.** 1963. Ben-Ami Lipetz. *Automation and Scientific Communication*, pp. 265–266 (see 7/64–151).

Objective evaluation of a documentation technique may be accomplished by introducing it in a limited test area and studying information user behavior in the test area versus control areas. Such an experiment is being devised to test the merit of the citation index as a reference aid for physicists (author).

1/65–170 **The Place of Indexing in the Design of Information Systems Tests.** 1963. A. J. Goldwyn. *Automation and Scientific Communication*, pp. 321–322 (see 7/64–151). Also available as report CSL:TR-3 of the Center for Documentation & Communication Research, Western Reserve Univ., Cleveland, Ohio. Aug. 1964.

Comparison of indexing techniques is difficult because current methodology distorts the role of indexing and its influence on total system design. Tests planned to evaluate indexing both as a variable and as a source of variation will hopefully lead to more realistic appraisals of comparative system performance (author).

1/65–171 **Index-Abstract Evaluation and Design.** April 1964. Al DeLucia. AMDO 15(2): 121–125.

1/65–172 **Cost Analysis and Simulation Procedures for the Evaluation of Large Information Systems.** April 1964. Charles P. Bourne and Donald F. Ford. AMDO 15(2): 142–149.

1/65–173 **Methodology for the Comparative Analysis of Information Storage and Retrieval Systems: A Critical Review.** July 1964. I. M. Klempner. AMDO 15(3): 210–216.

1/65–174 **Two Models for Retrieval System Design.** July 1964. D. J. Hillman. AMDO 15(3): 217–225.

Training in Documentation

1/65–175 **The Engineers Joint Council-Battelle Memorial Institute Coordinate Indexing and Abstracting Training Course.** Jan. 1964. John C. Costello, Jr. and Eugene Wall. JCHD 4(1): 26–29.

The Engineers Joint Council and its role are briefly discussed. Engineering information storage—retrieval—dissemination was identified as an area requiring EJC action. EJC's Information Systems Committee formulated the Engineers Joint Council Action Plan based on five fundamental considerations having

to do with: 1. the ability of engineers to recapture available technical information; 2. use of natural language as a fundamental method of communication among technical people; 3. providing a thesaurus as a "translation" medium with no restriction on choice of words by authors; 4. developing procedures compatible with the requirements of small individual "do-it-yourself" card files as well as large scale computerized IR systems; and 5. reducing the costs of duplicate abstracting and indexing of published articles. Progress to date toward meeting EJC's Action Plan targets includes:

1. Training journal editorial staffs to abstract and coordinate index with roles and links at time of publication.
2. Battelle's development of the EJC-Battelle training course in coordinate indexing and abstracting, and its regular presentation on a National basis. By utilizing the EJC abstracting and indexing approach on internally generated literature and on documents from non-EJC sources, individual organizations will be able to merge with their own work, the source-created abstracts and coordinate index terms with roles and links of up to 900 technical articles per month from EJC member society publications.
3. The EJC "Thesaurus of Engineering Terms," near publication and available soon.

A "users' conference" will be held by Battelle later in 1964 to exchange ideas, techniques and solutions to input and output problems and generally for discussion of operating problems and work (B.K.D.).

1/65–176 **Accrediting Information Science Programs.** 1963. E. J. Humeston, Jr., John F. Harvey. *Automation and Scientific Communication,* pp. 153–154 (see 7/64–151).

1/65–177 **Educational Requirements for Indexers in a Selective Dissemination System.** 1963. A. Resnick. *Automation and Scientific Communication,* pp. 163–164 (see 7/64–151).

An experiment was conducted in the SDI System at IBM-ASDD to determine if the education level of the encoder had a significant effect on the percentage of notifications regarded as relevant by the users. Considering all 800 documents encoded, high school educated encoders had a significantly higher percentage of their notices regarded as relevant than did four college educated encoders. Eliminating documents for which no notifications were sent, no significant difference was found between the high school and college level encoders (author).

1/65–178 **The Education and Training of Information Scientists in Biomedicine.** 1963. Isaac D. Welt. *Automation and Scientific Communication,* pp. 167–168 (see 7/64–151).

1/65–179 **Graduate Training in Information Science: Definitions and Developments at the Georgia Institute of Technology.** 1963. Dorothy Crosland. *Automation and Scientific Communication,* pp. 243–244 (see 7/64–151).

1/65–180 **Information Specialist Training Programs: Student Support: Graduate Prospects.** 1963. Theodore C. Hines. *Automation and Scientific Communication,* pp. 247–248 (see 7/64–151).

1/65–181 **An Important Need and Opportunity for A.D.I. Leadership in Information Science Education.** 1963. Robert L. Taylor. *Automation and Scientific Communication,* pp. 271–274 (see 7/64–151).

Current trends in information science education appear inadequate for the important need of the nation's practicing professional personnel for training in becoming information specialists or more proficient users of information systems. A particular educational program by A.D.I. is suggested to supplement others in meeting the presumptive need (author).

1/65–182 **The New Program for Training Information Technicians at the U.S. Department of Agriculture Graduate School.** 1963. John Sherrod. *Automation and Scientific Communication,* pp. 327–328 (see 7/64–151).

The Graduate School of the U.S. Department of Agriculture offers an organized program of study consisting of non-professional, library and documentation courses. These courses offer background information and training for the sub-professional assistant and others whose work requires a knowledge of information techniques. Individuals completing the program are granted a Certified Statement of Accomplishment in Library Techniques (author).

1/65–183 **Science Information Personnel: An Assessment and Projection.** 1963. W. K. Lowry. *Automation and Scientific Communication,* pp. 335–337 (see 7/64–151).

1/65–184 **The Education of Science Information Personnel —A Challenge to the Library Schools.** 1963. Alan M. Rees. *Automation and Scientific Communication,* pp. 343–346 (see 7/64–151).

1/65–185 **New Dimensions in Library Education: The Training of Science Information Personnel.** Oct. 1963. Alan M. Rees. SPLB, pp. 497–502.

Documentation – General

1/65–186 **The Role of "Letters" Journals in Primary Distribution of Information; A Survey of Authors of *Physical Review Letters*.** July 1964. Pauline Atherton. Amer. Inst. of Physics, New York, N. Y. 6 pp. Report AIP/DRP–64–1.

The authors of the 409 letters in the 1961 issues of *Physical Review Letters* were asked whether the material in their letters had been published subsequently in fuller and more detailed papers. Of the 377 who responded, 215 (57%) indicated that no further publication of the material had occurred. A study of the placement of abstracts of these letters in *Physics Abstracts* showed that 199 of the 409 letters were classified within the fields of Elementary Particles, Low Temperature Physics, and Magnetic Properties of Solids, with the rest distributed over all other fields. Work reported in 117 of these 199 letters was not published elsewhere. Of the 162 papers amplifying publication in letter form, over 70 were reported as published in *The Physical Review,* with the balance in 28 other journals and the proceedings of 15 conferences (author).

1/65–187 **Sixty Ideas in Sixty Months.** June 12, 1964. Lauren B. Doyle. System Development Corp., Santa Monica, Calif. 49 pp. Report TM–1952.

An account is given of the development of the author's ideas on the analysis and retrieval of information in the form of natural language text over a period of five years (July 1, 1958, to July 1, 1963). A month-by-month exposition is used to give time perspective as well as to avoid the over-emphasis of any particular era. Three charts show how the ideas relate to each other in evolutionary sequence.

The majority of the ideas reflect the development of the author's concept of the "statistical approach" in text analysis against the backdrop of the growth of research in statistical analysis of text throughout the nation. Smaller groups of related ideas come under topics such as automatic indexing, post-editing, evaluation of retrieval systems, differences in retrieval needs from one field to another, and the conduct of exploratory research (author).

1/65–188 **ASTIA Guidelines for Cataloging and Abstracting.** June 1962. Robert L. Murphy, Ed. Armed Services Technical Information Agency, Arlington, Va. 29 pp.

1/65–189 **Standard for Descriptive Cataloging of Government Scientific and Technical Reports.** Dec. 1963. Committee on Scientific Information, Federal Council for Science and Technology, Office of Science and Technology, Executive Office of the President, Washington, D. C. PB–181 605. 21 pp. OTS, $1.00.

1/65–190 **Writing Informative Titles for Technical Papers—A Guide to Authors.** 1963. R. A. Kennedy. *Automation and Scientific Communication*, pp. 133–134 (see 7/64–151).

Suggestions provided recently to authors at the Bell Telephone Laboratories on choosing good titles for their papers are presented (author).

1/65–191 **Applied Information Management System.** 1963. Elmer W. Heller. *Automation and Scientific Communication*, pp. 161–162 (see 7/64–151).

1/65–192 **Recent Activities in Foreign Exchanges.** 1963. Ines C. O'Brien. *Automation and Scientific Communication*, pp. 203–204 (see 7/64–151).

1/65–193 **Status Report on the Biological Sciences Communication Project (BSCP).** 1963. C. W. Shilling. *Automation and Scientific Communication*, pp. 205–206 (see 7/64–151).

Typical studies on the flow of life-sciences information by the Biological Sciences Communication Project deal with the citation of Russian literature, sources of support of scientific research, page-cost policy of journals, and the accuracy of titles in describing the contents of papers. Other programs are of interest to information administrators (author).

1/65–194 **Satisfying User Needs for Engineers Engaged in Planning, Design and Consulting Services for Industry.** 1963. Walter H. Rupp. *Automation and Scientific Communication*, pp. 231–232 (see 7/64–151).

A special machine-based IS&R system operated by experienced engineers for the benefit of user-engineers is described with emphasis on practical methods to satisfy user needs. The treatment is typical of some present and many industrial information systems envisioned in the future (author).

1/65–195 **Toward a Definition of Information Science.** 1963. L. B. Heilprin. *Automation and Scientific Communication*, pp. 239–241 (see 7/64–151).

An operational definition is attempted for the new composite discipline, "information science." The approach is based on the physics and psychology of messages. To include all steps in the "information transfer chain" the definition must remain general. A probable limit of specificity is suggested in the "final" definition offered (author).

1/65–196 **Documents and Information Exchange.** 1963. Bella E. Shachtman. *Automation and Scientific Communication*, pp. 257–258 (see 7/64–151).

The importance of document and information exchange is pointed out with an indication of its status and reference to some of the major exchange institutions. Suggestions for the future are decentralization of national responsibility for collecting, and the establishment of a national exchange information and advisory service (author).

1/65–197 **A Statistical Image Recognition Function Related to Integral Geometry and Its Instrumentation in a Statistical Transformation Generator.** 1963. George Tenery. *Automation and Scientific Communication*, pp. 267–268 (see 7/64–151).

The concept of using a statistical image transformation as a key step in image recognition has been studied by computer simulation, analytical studies relating the transformation to classical integral geometry, and construction of a prototype analog statistical image transformation generator (author).

1/65–198 **The Automated Multilevel Encyclopedia as a New Mode of Scientific Communication.** 1963. Herbert G. Bohnert and Manfred Kochen. *Automation and Scientific Communication*, pp. 269–270 (see 7/64–151).

The computer revolution encourages reexamination of the encyclopedia concept. No longer confined to book form, it can be enlarged to embrace several levels of condensation from the library-sized store, with updating, text preparation, access, and distribution handled differently at different levels, partially replacing, at base level, the journal system (author).

1/65–199 **The Reduction in Bulk Resulting from the Typesetting of Documents in Modern Composition Systems.** 1963. F. R. Kraft. *Automation and Scientific Communication*, pp. 275–276 (see 7/64–151).

A 43-page typewritten article occupied only 22 pages when automatically made-up by a computer and produced by a Linofilm phototypesetter. 2.6 pages of lists photo-offset directly from computer print out, or 3 pages of tables produced in like manner, could fit on only one typeset page without reduction in readability (author).

1/65–200 **Retrieval and the General Library.** 1963. Louis A. Schultheiss. *Automation and Scientific Communication*, pp. 283–284 (see 7/64–151).

1/65–201 **"Information Please" Services.** 1963. F. Ellis Kelsey. *Automation and Scientific Communication*, pp. 301–302 (see 7/64–151).

1/65–202 **Recording Text Information in Machine Form at the Time of Primary Publication.** 1963. Lawrence F. Buckland. *Automation and Scientific Communication*, pp. 309–310 (see 7/64–151).

It is proposed to develop a primary publication procedure which in addition to publishing the journal, records data needed for secondary publishing and storage and retrieval purposes. The limitations of typography and the requirement for a recording procedure which identifies content or item function are states. The problems of complex symbol representation are posed (author).

1/65–203 **Microprint as a Medium for Primary Publication.** 1963. Carlton M. Herman. *Automation and Scientific Communication*, pp. 311–312 (see 7/64–151).

Microprint has been demonstrated an acceptable medium for publication in an experiment conducted by the Wildlife Disease Association, with author composition costs slightly over seven cents per 3 x 5 inch card, up to 47 pages. Need is recognized for development of standards and improvement of accessory retrieval apparatus (author).

1/65–204 **The Use of Broadcast Television for Communication Among Scientists and Engineers.** 1963. John K. Mackenzie. *Automation and Scientific Communication*, pp. 315–316 (see 7/64–151).

The Science and Engineering Television Journal is an experiment in the use of broadcast television for inter-professional communication. Coordinated by the AAAS, under a NSF grant, 30 one-hour videotapes have been done in cooperation with 12 national societies. Evaluation indicates that the project is useful and a proposal is being considered for a second series with videotape distribution expanded on a national basis (author).

1/65–205 **The Relationship Between the Library and the Information Centers.** Feb. 1964. R. R. Dickison. *AEC Tech. Info. Bull.*, No. 10, pp. 14–15.

1/65–206 **The Heuristic Information Retrieval Game.** April 1964. Allen Kent. AMDO 15(2): 150–151.

1/65–207 **Effective Information—Searching Strategies without "Perfect" Indexing.** July 1964. R. J. Tritschler. AMDO 15(3): 179–184.

1/65–208 **The Changing Horizon of Information Storage and Retrieval.** June 1964. Harold Borko and Lauren B. Doyle. *American Behavioral Scientist* 7(10): 3–8.

1/65–209 **Information Retrieval: A Revolt Against Conventional Systems?** Feb. 1964. J. Mills. ASLP 16(2): 48–63.

1/65–210 **The Present State of Research into the Communication of Information.** Feb. 1964. B. C. Vickery. ASLP 16(2): 79–91.

1/65–211 **Science, Government, and Information.** Mar. 1964. Herbert Coblans. JDOC 20(1): 1–4.

Brief comment is given on the Weinberg Report. The central theme of the Report is that the scientific community must "improve its fundamental attitude towards scientific information and recognize . . . that the handling of technical information is a worthy and integral part of science." A higher status should be granted to the compiler of scholarly reviews and critical bibliographies and to the documentalist. While he feels that government should support specialized information centers, the author would go a step further and suggest that the large government agencies could greatly help and speed up international standardization by working through the professional scientific unions. Concerning the mission-discipline duality, the enormous American mission systems for information can become very wasteful of manpower and money if their work is not properly coordinated. Since the same tendencies exist in European countries and at the regional level, this is an area for international cooperation through the specialized agencies of the United Nations (D.C.W).

1/65–212 **The Coverage of Heavy Electrical Engineering Periodical Literature by Abstracts Journals.** June 1964. P. Clague. JDOC 20(2): 70–75. Tables.

The study investigated the proportion of articles abstracted and the length of time before abstract publication and indexing for sixteen journals. The abstract services used were the *Electrical Engineering Abstracts, Engineering Index, CEGB Digest,* and *ERA Weekly Abstracts.* All four services together covered only 85 percent of the articles checked. The typical time required for abstracting was twelve to eighteen months (D.C.W.).

1/65–213 **Problems Involved in the Graphic Reproduction of Mixed Scripts.** June 1964. J. T. Harrison. JDOC 20(2): 76–86.

Publications combining English and exotic scripts face high typewriting costs and they usually have limited sales so the unit cost is also high. Problems and pitfalls of such composition are discussed; e.g., required size of fonts, direction of setting, length of ascenders and descenders, and hand composition adjustments. Easier preparation is obtained through use of photo-composing machines or typewriters with interchangeable type bars or type sets. Still, the painstaking hand assembly of different typesettings is required for the intermixing of non-Roman alphabets. Problems here are also mentioned (D.C.W.).

1/65–214 **Chemical Writing.** 1963. M. G. Mellon. JCHD v(1): 1–5.

The current low standard of writing in published literature in chemistry and chemical technology, and by extension to technical writing in general, is criticized and remedies are suggested. Published literature is divided into primary sources containing new information; e.g., periodicals, government bulletins, patents, dissertations and manufacturers' technical bulletins; and secondary sources which contain primary source materials arranged for particular purposes; e.g., abstracting journals, index and review serials, bibliographies, reference and textbooks, etc. Five means by which poor literature can be improved are: better use of English language, effective use of formulae and equations, effective tabular presentation of numerical data, satisfactory drawings and graphs, and appropriate pictures. General requirements of good technical writing are: accuracy, clarity, brevity, consistency and interest value of subject. Recommended writing procedure is: collection and arrangement of material, determination of writing form and best means of expression, production rough draft followed by final revised manuscript. There is a one page bibliography of selected references (V.E.Y.).

1/65–215 **The Charter: A "Must" for Effective Information System Planning and Design.** Jan. 1964. John C. Costello, Jr. JCHD 4(1): 12–20.

A set of principles basic to the sound planning, design, development, and operation of information systems is presented. Several topics to be considered in drawing up a charter are discussed. Although the charter provisions proposed in this paper may be more or less comprehensive than required in a given situation, a well developed charter is recommended as an excellent control device for system efforts. Under the heading, *General Principles,* the author tells what the charter must accomplish and why. In addition, thirteen areas of consideration are discussed: User Group, Coverage, Service, Organization, Personnel, Physical Facilities, Document Storage, Control, Quality Control, Scheduling, Mechanization, Compatibility, and Adaptability. The charter as a flexible framework with provision for both system and charter revision is emphasized. The charter must reflect higher management's attitude toward the system. Provisions should be included for planned system adjustments in times of financial fluctuations. This paper, which presents a composite picture of observed systems, can help reduce duplication and errors in information system planning (B.K.D.).

1/65–216 **Randomized Binary Seaching with Tree Structures.** March 1964. Harry A. Clampett, Jr. JACM 7(3): 163–165.

A more efficient method of using tree structures is proposed, using both plus and minus branches in the search path. Very significant gains result when the search key includes alphabetic characters (author).

1/65–217 **Technical Services in 1963.** Spring 1964. Maurice F. Tauber. LRTS 8(2): 101–111.

Developments marked an effort to relate the field of documentation more closely to what has been described as traditional librarianship. Notable are joint committees of ADI and SLA, the ALA Interdivisional Committee on Documentation, conferences, and new curricula in a half dozen library schools. Other specific publications and meetings related to documentation are described (D.C.W.).

1/65–218 **European Report: Reprography Congress, Document Reproduction Activities in French and British Libraries.** Spring 1964. Allen B. Veaner. LRTS 8(2): 199–204.

Gives brief report on the First International Congress on Reprography, in Cologne, which attracted ten thousand visitors and delegates. It was held concurrently with the International Trade Exhibition for Reprography at which nearly one hundred firms exhibited. Over a hundred papers were presented with simultaneous translation service; publication in abbreviated form will be by Verlag Dr. Othmar Helwich. Seven of the best papers are cited, including "Reprographic Methods in Data Processing" by

Ronald H. Kay and "Microfilm as a Systems Tool" by Ernest P. Taubes. Basic equipment and techniques are mentioned as exist in six photoduplication laboratories in France and six in England. For instance, the British Museum Map Room is developing a system using aperture cards which will offer an efficient, simple, and economical method for the exchange of sheet maps (D.C.W.).

1/65–219 **The Changing Character of the Catalog in America.** Jan. 1964. David C. Weber. *Lib. Quarterly* 34(1): 20–33.

Describes the development of forms of the library catalog in the United States from the seventeenth century to the present. Consideration of the use of manuscript and printed book catalogs is followed by treatment of the card form. Coverage also includes developments since 1951 in printed catalogs, in microformats, and data processing equipment. The advantages and limitations of these techniques are explained. The conclusion is reached that success in information retrieval is not dependent on the form of the instrument, yet the form will have a marked effect on convenience and speed. These are conditioned by factors such as requirement for portability, need for copies in several locations, number of simultaneous users to be accommodated, importance of currency, requirements for flexibility in intercollating new entries, the importance of browsing, the importance of cost, and the degree to which the collection changes through additions and deletions (D.C.W.).

1/65–220 **The Documentation of Science Organization as an Emerging New Branch of Scientific Information.** 1962. György Rózsa. *Magyar könyvszemle* 78, No. 4, pp. 265–277 (Hungarian Journal. Available at LC).

1/65–221 **Technical Reports on Microfiche—What Price Unitization?** April 1964. William R. Hawken. NAMN No. 69, pp. 195–207.

1/65–222 **Some Problems of Scholarly Publication in Photocopy.** April 1964. Robert B. Eckles. NAMN No. 69, pp. 215–218.

1/65–223 **Committee on the Utilization of Stored Data Systems.** 1964. *The Professional Geographer* XVI(4): 36–39.

Committee report is concerned with two primary aspects of *data banks* [sources]: 1. identification and listing of those presently available; 2. making of new data banks which take the geographer into consideration when data are recorded. Report is divided into five parts: inventory of data already stored, problems encountered in stored data use, nature and format of data stored in future, costs of storing and distributing data, and disclosure problems (J.J.C.).

1/65–224 **Geographic Ordering of Information: New Opportunities.** 1964. Brian J. L. Berry, Richard L. Morrill, and Waldo R. Tobler. *The Professional Geographer* XVI(4): 39–44.

Report of the Geographic Coding Subcommittee, Census Advisory Committee, Association of American Geographers; covering the collection, storage, retrieval, and analysis of data required by academic geographers. Discusses six potential benefits from modern data systems: mapping by computer, partitioning and aggregation, merging of data systems, transformation of geographic spaces, descriptive statistics, and intermediate analysis. Also considers problems yet requiring solution (J.J.C.).

1/65–224 **The Changing Character of the Catalog in America.**

1/65–225 **Information Retrieval.** May 10, 1963. Ralph R. Shaw. SCIE Vol. 140, No. 3567, pp. 606–609.

1/65–226 **Primary Scientific Publication and the Federal Government.** May 10, 1963. Burton W. Adkinson. SCIE Vol. 140, No. 3567, pp. 613–617.

1/65–227 **A Working Microfilm Card Catalog.** March 1964. H. Matsumiya and M. Bloomfield. SPLB 55(3): 157–159.

1/65–228 **Bibliography on Reproduction of Documentary Information.** March 1964. Loretta J. Kiersky. SPLB 55(3): 160–165.

1/65–229 **The Information Goals of Engineers Joint Council.** March 1964. Stanley Klein. SPLB 55(3): 143–147.

1/65–230 **New Bottles for Old Wine: Retrieval and Librarianship.** May 1964. Alan M. Rees. WLDU, pp. 773–779.

1/65–231 **Book Catalogs.** 1964. Robert E. Kingery and Maurice F. Tauber. Scarecrow Press, Inc., New York, 330 pp. $7.00.

1/65–232 **Readings in Special Librarianship.** 1964. Harold S. Sharp. Scarecrow Press, Inc., New York, 714 pp. $15.00.

1/65–233 **A Conceptual Framework for the Augmentation of Man's Intellect.** 1963. Douglas C. Engelbart. *Vistas in Information Handling,* pp. 1–29 (see 7/64–150).

Man is confronted by increasingly complex problem situations which require use of great stores of data and experience together with artifacts of simplest to most sophisticated electronic aids. His intellect, then, must be extended so that comprehension is better and more quickly achieved with concomitant reaching of solutions to these complex problem situations. This opening chapter provides a conceptual framework for the augmentation of man's intellect (author).

1/65–234 **Patent Examination and the Need for Research.** 1963. Richard A. Spencer. *Proc. of the Social Statistics Section.* Amer. Stat. Assoc., Washington, D. C., pp. 85–90.

1/65–235 **Application of Information Retrieval Principles to Education.** 1963. Frederick L. Goodman. *Automation and Scientific Communication,* pp. 51–52 (see 7/64–151).

1/65–236 **Matching of Descriptors in a Selective Dissemination System.** 1963. Jack W. Hilf. *Automation and Scientific Communication,* pp. 65–66 (see 7/64–151).

1/65–237 **Author Extracts in Addition to Technical Reports: An Aid Toward Timely Dissemination of Technical Reports.** 1963. Julius Frome, Joseph F. Caponio, and Virgil Brown. *Automation and Scientific Communication,* p. 73 (see 7/64–151).

The essential and salient results contained in engineering reports are being repackaged in an extract easily digested and assimilated by the researcher in the field. These extracts not only save the time of the researcher and eliminate or reduce reproduction workload, they provide wanted information in a form more readily usable by different audiences (authors).

1/65–238 **Direct Documentation, Analogue Documentation, and Information Retrieval.** Jan. 1964. Robert F. E. Schweizer. AMDO 15(1): 23–25.

1/65–239 **Research Approach to University Library Problems.** May 1963. Robert H. Muller. *Coll. and Res. Libs.,* 24(3): 199–203.

The Research and Development Committee of the University Libraries section of ACRL, believing that more research into library problems should be carried out in order to provide a sounder basis for policy determination, has invited a number of librarians to suggest topics worthy of systematic investigation. Standards of output in cataloging and acquisition: Felix Reichmann. It will be necessary to analyze the technical operations and

to calculate and test the length of performance of operations. Measuring library use: R. T. Esterquest. The usefulness of a library cannot be measured by simple reference and load statistics. Suggestions include: measuring the use made of journals and counting open-stack browsers and the duration of their browsing. Success and failure in library use: Andrew Eaton. Analysis should focus on individual students to discover degree of success, variations among students, and the factors associated with success and failure. Attitudes of university administrators toward libraries: John F. Harvey. This would aim to discover administrators' evaluations of libraries and librarians' evaluations of administrators. (*Lib. Sci. Abstracts*).

1/65–240 **Searching Techniques in the Literature of the Sciences.** May 1963. Hugh E. Voress. *Coll. and Res. Libs.* 24(3): 209–212.

Describes the procedure for searching technical literature at the U.S. Atomic Energy Commission's Division of Technical Information Extension, Oak Ridge, Tennessee. The searcher must determine the exact needs of the requester, the scope of the subject, the sources to be checked, subject headings to be examined, and the time period to be covered. References and abstracts must be evaluated and author indexes checked. All references must be edited into a single style. Any search undertaken by the Division which has taken appreciable time, or which has resulted in a significant list of references, is published and distributed to AEC contractors and depository libraries. All bibliographies and searches are listed in the TID–3700 series of bibliographies of atomic energy literature in the cumulation, TID–3043 (Rev. 2), Bibliographies of Interest to the Atomic Energy Program. (*Lib. Sci. Abstracts*).

1/65–241 **Library Statistics of Colleges and Universities, 1961-1962.** May 1963. *Coll. and Res. Libs.* 24(3): 235–238.

1/65–242 **Information Retrieval: The Man-Machine Interface.** Nov. 1963. A. J. Goldwyn. *Data Processing for Education* 2(11): 1–10.

1/65–243 **Developing Advanced Business Information Systems.** Oct. 1963. Edward L. Weinthaler, Jr. *Data Processing for Management* 5: 9–14.

The management scientist is playing a key role in the design, development, and installation of information systems. The design objective is to build a system which will facilitate the decision making process in production, marketing, and man-power (*Battelle Tech. Rev.*).

1/65–244 **Information as an Instrument for the Management of an Enterprise.** 1962. Z. Jarkiewicz. *Ekon. i organiz pracy* (1): 24–26. In Russian.

1/65–245 **Periodicals and the Three Card System.** April 1963. T. R. Vedachalam. *Herald of Library Science* 2(2): 74–81.

Stresses the importance of periodical publications in libraries and describes and criticizes the one card system for recording their receipt. The front of the card is ruled so that issues can be marked off on receipt and the reverse indicates the payment of accounts. Dr. Rnaganathan introduced the three card system at Madras U.L. in 1930 and these cards are described and illustrated. The first card is the register card and these are arranged in alphabetical order, issue being marked off on receipt. The second card is the check card and these are arranged to show the month and week when the next issue is due. This allows non-receipt of an issue to be noticed immediately. The third, or classified index card, is placed in a classified sequence to show which periodicals are available on a particular subject. The advantages of the system are enumerated (*Lib. Sci. Abstracts*).

1/65–246 **Who'll Tame Information Flood?** Jan. 16, 1964. Gerald L. Phillippe. *Iron Age* 193(3): 53.

The president of General Electric Co. states that financial managers must identify the kinds of information needed in managing the business; produce the mass of business data needed; develop better means for processing and transmitting business data; interpret what the data mean; and develop improved approaches to budgeting, forecasting, performance standards, and measurement of the whole company (*Battelle Tech. Rev.*).

1/65–247 **The Balanced Tree and Its Utilization in Information Retrieval.** Dec. 1963. Walter I. Landauer. IREL EC–12(5): 863–871.

To translate descriptors into memory locations a memory organization scheme called the balanced tree is introduced. The descriptors that describe the information to be stored or retrieved constitute quasi-inputs to the tree while the outputs are lists on which the information identified by the descriptors is stored. The balanced tree thus provides a strategem to effect a fast information retrieval with a limited amount of serialized scanning. The algorithms for storing in and retrieving from the balanced tree are outlined. While in a randomly growing tree, the shape of the tree depends on the order of the input, the balanced tree is independent of this order. The expected number of rearrangement steps to keep the tree balanced was derived from combinatorial considerations. Numerical results were obtained by machine computations and are presented in this paper (author).

1/65–248 **Literature Research for a Space Materials Research Program.** Oct. 1963. E. G. Kendall, Edythe Moore, and Charles Hays. JCHD)(4): 232–234.

Describes a literature search on refractory materials at the Materials Sciences Laboratory of the Aerospace Corporation during 1962 (*Battelle Tech. Rev.*).

1/65–249 **Book Publishers' Interests in Reprographic Copyright.** Aug. 1963. Curtis G. Benjamin. LIBJ 88(14): 2837–2841.

An investigation, sponsored by the National Science Foundation, into an allegation that the Copyright Law was stifling scientific progress, concluded that the law was both non-restrictive to science and technology and necessary for the protection of copyrights. The book industry's position on certain proposed changes in the U.S. copyright law as it affects "fair use" and the photocopying of an entire publication is made clear by reference to statements made by the industry's Joint Copyright Committee. Publishers are convinced that in many cases a book is reproduced as a substitute for a copy that might be purchased, and a further reduction in photocopying charges might lead to more copying of this kind. This could have serious adverse effects on the publication of advanced treatises and monographs which are dependent for their basic support on the library market. The legal and economic principles are the same in the case of mechanized storage and retrieval systems which supply copies of copyright material, but such a system could have a built-in subsystem for calculating fees due for the use of such material. The question of payment for the right to reproduce and use copyright material has not been seriously tackled by librarians, whose attitude has been that this is a problem for the publisher to solve. With the prospect of cheaper copying methods, some of the savings could be passed on to copyright holders in the form of higher fees, or a Foundation might provide a subsidy to start and administer a system for fee collection. Authors and publishers would be satisfied with nominal reproduction fees if they were not also expected to pay the costs of establishing and operating the system for collecting them (*Lib. Sci. Abstracts*).

1/65–250 **Cataloging Problems in Medical Libraries.** Spring 1963. Wilhelm Moll. LRTS, pp. 197–199.

1/65–251 **Cataloging Small Manuscript Collections.** Summer 1963. Michael Jasenas. *Lib. Resources* 7(3): 264–273.

1/65–252 **Cataloging and Classification in Junior College Libraries.** Summer 1963. Arthur Ray Rowland. LRTS 7(3): 254–258.

1/65-253 **A Square Inch for Libraries.** Summer 1963. Paul E. Vesenyi. LRTS 7(3): 294–296.

Some standardization is necessary to simplify the handling of current copies of periodicals in even the smallest library, particularly as to where details of date, volume, etc., appear on each copy. All publishers should agree: 1. to reserve a square inch on the front or back cover of the periodical for identification of the particular issue—title, volume, issue numbers, date; 2. the assigned spot must be on a previously agreed position on every periodical. The immediate advantages would be time-saving and reliable identification, faster shelving and retrieving, easier processing for the bindery, the possibility of standardization of serials' titles, and simplification of bibliographical citation. A future advantage would be the ease of codification for data processing equipment (*Lib. Sci. Abstracts*).

1/65-254 **The Analysis of Structural Elements of Reference Functions in University Libraries.** 1963. Sakari Ide. *Library Science* (Japan) (1): 21–32.

Reference service and reference work are compared by examination of Rothstein's theory and the ALA's descriptive list of professional and non-professional duties. Personal assistance has four phases: 1. to indicate the location of library material sought by a reader or to guide him to the right department; 2. to assist a reader to approach primary sources through secondary sources; 3. to advise a reader on what to read for the information he needs; 4. to offer a reader the information he needs. The last, called information service, is discussed in detail as the most important aspect. The structural elements of reference functions as dealt with by Shores, Christoffers, and Wilson & Tauber are outlined, and Wyer's and Lyle's classifications by groups to serve are reviewed. Adopting Lyle's classification, the writer combines two functions (action, preparation) with Shore's six functions (supervision, information, guidance, instruction, bibliography, appraisal) (author).

1/65-255 **Aspects of the Financial Administration of (Academic) Libraries.** April 1963. LIBT 11(4).

Outside funding of academic libraries. David Kaser 353–361. Private benefactors provide 7% of funds of large libraries, but more often now gifts are channeled through foundations and are for capital expenditure. Federal support is assuming an increasing role as research libraries are recognized as a vast national resource essential for information needs worthy of direct aid; important indirect Federal support includes document depositing and technical reports center. Increasing numbers of large capital grants are made by the National Science Foundation. Disadvantages of outside funding include: 1. influence, if only negative, on academic programs; 2. necessity of political expediency; 3. grants for exotic schemes rather than improvement of work already being done, although funds may sometimes be re-directed. Chief librarians must be mainly fund-raisers with a dynamic public relations program to attract grants from all possible sources.

Operating costs of college and university libraries. Ralph H. Parker. 376–383. Inflation and the greater number of students following higher degrees are rapidly increasing the library's annual cost per student. Librarians have greater special service demands because of large research projects and salaries expenditure is more than double that on library materials. Even greatly increased funds cannot meet rising costs; attempts at reduction include elimination of services and/or mechanization, especially in charging and cataloging procedures. Simplified transaction number control systems and punched card call slips are used; the extensive post-war shift to open stacks and limitation of paging services has an economy as well as an educational function. Limited cataloging is used for arrears and material of small probable use (LC is studying shelf classification for research libraries) or LC cards are used without modification. Mechanization is still very limited, but the University of Illinois study of complete automation of records has a broad usefulness for universities. Smaller colleges are realizing the need for joint acquisition and cataloging to use technical advances.

Academic budgets and their administration. 1962. James H. Richards, Jr. 415–426. Major issues are: 1. development of special library facilities; 2. optimum size of collection. Unprecedented growth and complexity have meant reorganization of higher educational administration and realistic long range planning, though the need to conform to the general institutional pattern has prevented adoption of performance budgeting. Library budgeting is generally recognized as a special field where the librarian has discretionary powers, the library committee having an advisory and liaison function. Increased student population demands self-surveys, plans for new facilities, and major fund-raising programs (*Lib. Sci. Abstracts*).

1/65-256 **Changes in the Concept of "Scientific Literature."** 1963. (Editorial.) *Mental Health Book Review Index* 8(13). Reprint.

1/65-257 **Military Specifications: Photographing of Construction/Architectural Drawings, Maps, and Related Documents, 105 mm; Requirements For.** Dec. 1, 1962. MIL-P-9879A. Department of Defense Reprinted in NAMN, (65): 16–24. August 1963.

1/65-258 **Superpositioning of Standard Punched Cards for Information Retrieval.** 1962. G. E. Vleduc. *Naucnotechniceskaja informacija* (5): 25–30.

1/65-259 **First Aid—The Literature of Non-Destructive Testing.** May-June 1963. G. M. Corney. *Non-destructive Testing.* (21): 167–174.

The present volume of technical information and problems of finding specific literature are discussed with particular attention to the difficulties of finding papers on non-destructive testing. Main sources, specific engineering journals, are described, together with ways of tracing official and non-government reports and patents. Abstract journals and indexes, particularly in technology and metallurgy, are described with notes on literature searching techniques. Methods of maintaining indexes are reviewed and the relative advantages of manual versus mechanized files are debated in some detail (*Lib. Sci. Abstracts*).

1/65-260 **Documentation: Handmaiden to Science.** 16 Dec. 1963. Rowena W. Swanson. *OAR Research Review* 11: 12–14.

1/65-261 **Book Order—Horse and Buggy or Jet Age?** Feb. 1963. Mary Shortt. *Ontario Lib. Rev.* 47(1): 5–8.

During the past half century practically everything has changed in Ontario except the way public libraries order their books. Librarians still pick books from publishers' lists, still do their own purchasing alone, and still have not found a way to pool their experience in selecting books. In 1961, the catalogers of the Metro Toronto public libraries made a six months' survey of the non-fiction titles catalogued for the first time in each of the 13 systems. Miss Shortt writes that the results of this survey indicate that present selection methods can nullify the efforts of even the best bookmen. She goes on to suggest essential reforms. The chief of these is the setting up of a central order department by several library authorities to give a system of cooperative book selection. The system is described and the writer mentions and deals with some anticipated objections (*Lib. Sci. Abstracts*).

1/65-262 **The Publication Charge Plan in Physics Journals.** June 1963. *Physics Today* 16(6): 45–57.

For more than 30 years, research institutions have shared part of the expense of publishing the results of physics research they have sponsored. This article summarises the plan from its inception through to recent developments. The record of page charge rates and an analysis of page charge honoring are also given.

1/65-263 **PNBC: Static or Dynamic.** July 1963. Desmond Taylor. *PNLA Q.*, 27(4): 208–213.

Proposes that the Pacific Northwest Bibliographic Center become a model for a demonstration of automation techniques on a bibliographic center. There are three catalogs in the center, the LC Catalog on cards, the PNBC Union Catalog consisting of 3,675 drawers, and 450 drawers of cards waiting to be filed into the Union Catalog. Most of the work done consists of routine clerical operation. The University of Washington Computer Research Laboratory has its own equipment which could be rented more cheaply by the University's School of Librarianship which would add prestige to it by conducting the research. The subject approach could be greatly extended with automated techniques and there would be a great saving in the production of catalog cards by using book catalogs. The scheme could be financed by subscription, based on library budgets. Outstanding special collections such as the Boeing Special Library and the Washington State Historical Society could also be added to the forty-two libraries which make up the present Union Catalog (*Lib. Sci. Abstracts*).

1/65–264 **Documentation and Its Problems.** Nov. 1963. S. R. Ranganathan. REDO 30(4): 127–128.

Using the metaphor of umbra and penumbra, the author defines the tasks of documentation service (umbra)—in terms of his 5 laws of library science—and the documentation process (penumbra). For the latter, two categories of operations are distinguished; those which are part of the library profession and others (translating, reprography, machine retrieval, and translation). In the first category, classification is considered the most pervasive. Principles for a faceted depth classification are outlined in view of newly emerging thought. Subjects needing research are indicated (author).

1/65–265 **Design of Books by Typewriter/Offset.** Feb. 1964. Julian Rollnick. REDO. 31(1): 20–24.

With the appearance of small offset litho presses and variable typewriters, it has become possible to publish literature of limited demand at reasonable cost. Following step by step the process of preparing manuscripts for offset printing, directions are given for the non-professional publisher (author).

1/65–266 **Challenge to Editors of Scientific Journals.** Sept. 13, 1963. J. R. Porter. SCIE. 141(3585): 1014–1017.

1/65–267 **Implications of the Copyright Law on the Dissemination of Scientific and Technical Information.** Nov. 1963. John C. Koepke. SPLB. 54(9): 553–556.

George Fry & Associates conducted an analysis of the influence of the copyright law for the National Science Foundation. The goals, methods, and results of this study are given. BATR.

1/65–268 **Organization of Scientific and Technical Information in the USSR.** 1962. D. Isakovic. *Technika* (5): 113–118. In Russian.

1/65–269 **Photographic Reproduction and Copyright.** July-Aug. 1963. Bibliotheque Nationale, Paris. UNDL. 17(4): 224–241.

Results of survey undertaken on behalf of UNESCO, preceded by a summary of the questionnaire and notes on its distribution to 335 bodies in twenty countries and to six international organisations. One hundred sixty-four replies were received and these are analyzed here. Certain similarities are found between establishments of the same type, but although all the establishments are concerned with protection of copyright and, although they apply the principle of protection in different ways, they all apply it with sufficient flexibility to avoid hampering scientific work: the differences may be attributed to national legislation, the nature of the documents in question, or the nature of the establishment. Certain of the results obtained are presented in tabular form (*Lib. Sci. Abstracts*).

1/65–270 **On the Improvement of the Organization of Scientific and Technical Information.** 1962. *Vestnik AN SSSR* (9): 121–122.

1/65–271 **Current Problems of Scientific and Technical Information.** 1962. A. I. Mihajlov. In symp. *Vycisl. i inform. technika,* pp. 6–12.

1/65–272 **Possibilities and Limitations of Modern Documentation Methods in Clinical Medicine.** Gustav Wagner. 10 Aug. 1963. West Germany. *Aerztliche Forschung* 17(8): 432–440. In German.

The traditional documentation of clinical findings is urgently in need of revision. To rationalize clinical documentation we have now the techniques of mechanical selection. Compared with manual methods, the automatic punched card methods are far superior in many respects. The most essential points which, according to our own experience, should be watched in introducing automatic punched card methods in clinical work are briefly discussed. One of the first suggestions for a coordinated documentation of health reports with modern means is the "General Health Report Heading" (*Allgemeiner Krankenblattkopf*), recommended by the medical committee of the German Society for Documentation and meanwhile adopted by many German hospitals. The limits of clinical documentation of findings are not confined to machines but rather to the frequent insufficiency of medical information which concerns most all the concepts of disease. The improvement of nosological systematics therefore is a most urgent problem. The general introduction of automatic methods of documentation in hospitals seems only a question of time. The medical faculties of the universities should heed this development through the creation of institutes for medical documentation and statistics as well as corresponding chairs (author).

1/65–273 **Repackaging of Scientific and Technical Information.** Oct. 1963. Julius Frome and Joseph F. Caponio. JCHD, 3(4): 229–232.

The authors drawing on their experience at DDC describe a specialized information package designed to meet the needs of the scientist, engineer, or technical administrator in the acquisition and location of scientific information pertinent to his field of work. It proposes an information service system combined into one publication and its attendant services, the past, current, and projected work efforts on a specialized area in science or technology. The principal elements of this new service are: a custom bibliography (with abstracts), an index to the bibliography section using current descriptive terms used in the area of the discipline or technology being covered; a listing containing information on current research projects; and a supplementary detailed vocabulary section listing terms by which detailed information can be retrieved, with some specific guidelines for obtaining the desired information quickly. The authors go into detail on the four principal elements of the new service package (J.W.M.).

LITERATURE NOTES—Please enter my subscription for one year commencing with the January 1965 issue. At $4.00 per year, payment is enclosed ☐

bill me. ☐

Name__________ Title__________

Address ☐ Business ☐ Home__________

City__________ State__________ Zip__________

Your Firm Name__________

Literature Notes Separate

The Literature Notes section of American Documentation is now available as a separate publication.

Literature Notes will continue to be an important section of American Documentation. However, popular demand has persuaded ADI to offer this separate quarterly subscription as an offprint.

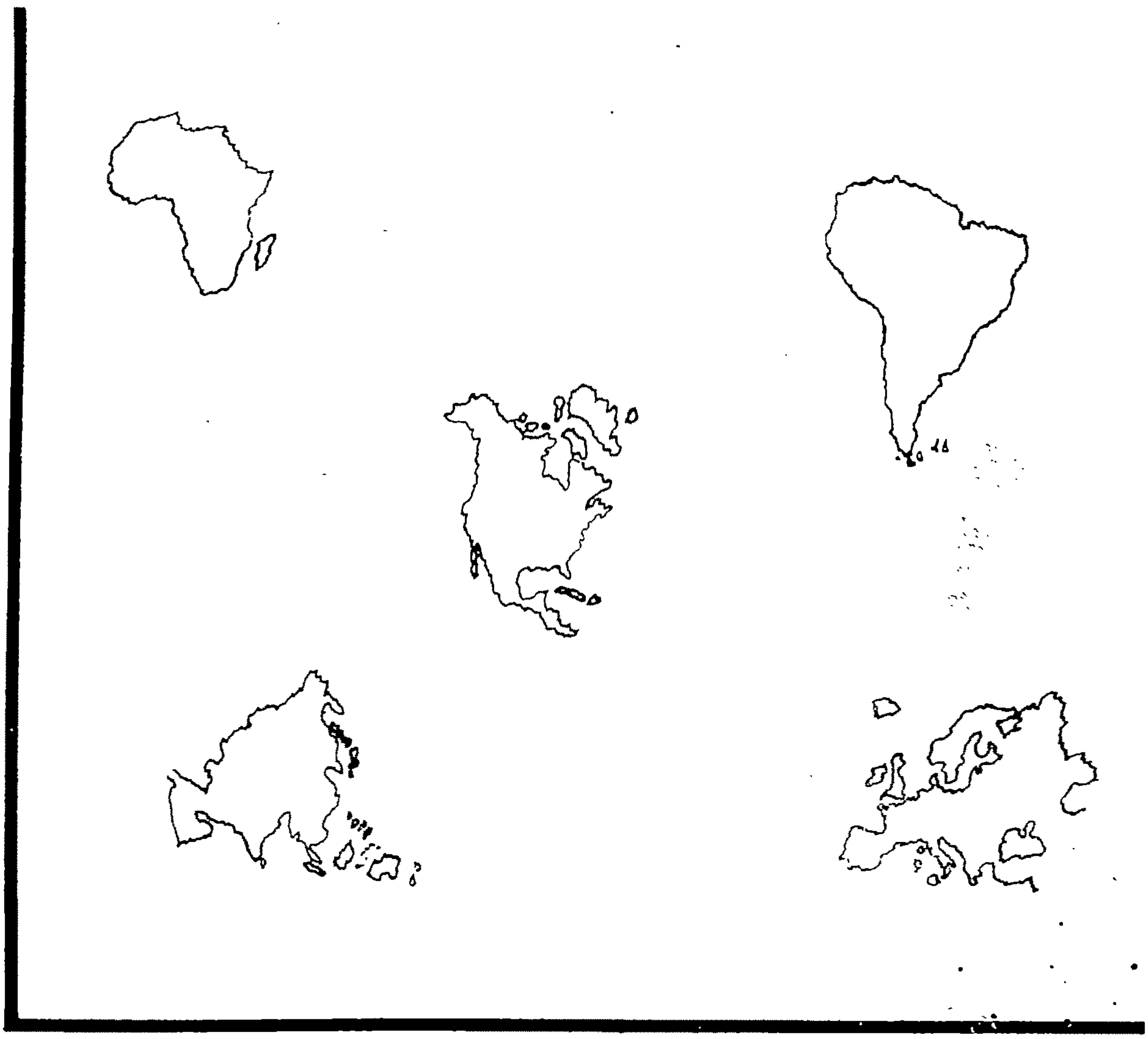
On a Calendar year basis, the subscription price of Literature Notes is $4.00 per year, commencing with the January 1965 issue.

Return the coupon above. Payment with the order is requested.

American Documentation

APRIL
1965
Vol. 16, No. 2

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE



AMERICAN DOCUMENTATION

American Documentation is a publication of the American Documentation Institute. Its purpose is to be a scholarly journal in the various fields of documentation and to serve as a forum of discussion and experimentation.

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), one copy in the institutional membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

Instructions to authors are given in Volume 14, No. 2 (April 1963). All manuscripts, brief communications (1000 words or less), announcements, or letters should be sent to:

Mr. Arthur W. Elias, Editor
American Documentation
Institute for Scientific Information
325 Chestnut Street
Philadelphia, Pennsylvania 19106

Correspondence on reviews should be addressed to:

Dr. Theodore C. Hines, Review Editor
American Documentation
54 North Drive
East Brunswick, New Jersey

Correspondence on Literature Notes should be addressed to:

Mr. Charles P. Bourne, Literature Notes Editor
American Documentation
Stanford Research Institute
Menlo Park, California

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts,* and *Science Citation Index.*

American Documentation is entered for second class mailing at Washington, D. C.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 16, No. 2 **APRIL 1965**

Published quarterly by AMERICAN DOCUMENTATION INSTITUTE and printed by Spartan Books, Inc., 1250 Connecticut Avenue, Washington, D. C. Second class postage paid at Washington, D. C.

Special Message

ANNOUNCING THE *ANNUAL REVIEW OF INFORMATION SCIENCE*

On behalf of the American Documentation Institute, I am pleased to announce the receipt of a grant from the National Science Foundation to initiate a new publication: *Annual Review of Information Science*.

As stated in the letter announcing the award, "The proposed *Review* will provide a much needed consolidation, on a regular and systematic basis, of the vast and growing body of research, development and applications work now being reported in this field."

The grant ($60,500 for two years) covers all costs up to but excluding those of publication. Already, several publishers have expressed interest. It is hoped that the publication, given the initial help of the National Science Foundation, will become self-supporting.

To implement the project, a subcontract is to be drawn between the American Documentation Institute and the System Development Corporation whereby SDC will collaborate with ADI by supplying an editor and staff. The Editor will be Dr. Carlos A. Cuadra, who has been instrumental in organizing the joint project. To assist ADI and the Editor in their joint task and to help maintain a high quality in content, coverage and writing, I have appointed an Advisory Committee for the *Annual Review of Information Science*, consisting of the following ADI members:

Dr. Ruth M. Davis, *Chairman*, Department of Defense
Charles P. Bourne, Stanford Research Institutes
Arthur W. Elias, Institute for Scientific Information
Dr. Robert A. Fairthorne, Herner and Company
John Markus, McGraw-Hill, Inc.

This grant represents an outstanding opportunity for ADI to serve information science and the scientific world. Its fulfillment will take much hard work on the part of all concerned—the Editor, the Advisory Committee, and the specialists who, in various branches of information science, will each monitor the literature and write a definitive chapter summarizing annual advances in that branch. An opening chapter may attempt to give a brief résumé of these summaries.

Some question may arise in your minds about the relation of the proposed *Annual Review* to our already established annual Report of the Committee on Organization of Information (COI). As I see it, the coverage of the *Review* and COI Report will be equally wide, but the approach differs. The COI Report attempts to forecast trends in IS as observed by its members. No attempt is made to cover the literature or to summarize it. Further, *opinion* on trends is given by the COI members, and the topics are not limited to given areas. The aim is a collection of birds' eye views by a group of experienced observers, as contrasted with a collection of careful, scholarly views, each of a part of the field. For the present, therefore, the COI Report will continue to fulfill its somewhat different need. It is significant, however, that at least two ADI members will contribute to both the *Annual Review* and the COI Report.

Because of the time and effort necessary to start a new scientific scholarly publication, it is expected that the first volume will appear in early or mid-1966. It will *review* 1965 and *some* earlier work. Thereafter, every attempt will be made to issue the *Review* as close to the end of the calendar year as possible, i.e., to combine timely coverage with highest quality presentation.

L. Heilprin, *President*
American Documentation Institute

Guest Editorial*

Almost seven years ago, the first International Conference on Scientific Information in the U.S.A. (Washington I) proved to be one of those obvious landmarks in a decade of turbulent activity. In a few months, Washington will once again be host to an international meeting, this time the 1965 Congress of the F.I.D. In taking stock, we are immediately struck by two significant developments: one national, the other international—the new strength of the A.D.I. and the growing impact of U.S. participation in international documentation.

A closer look at the past few years shows the thorough involvement of such large mission-oriented agencies as the A.E.C. and N.A.S.A. in documentation—a commitment which has had great repercussions in Europe and the rest of the world. Wherever there has been research and development in nuclear energy, *Nuclear Science Abstracts* has contributed towards setting the pattern. Especially in Europe, EURATOM, and more recently, E.S.R.O., are shaping their documentation policies to utilize the stored capacity on magnetic tapes that are gradually becoming available from across the Atlantic. This probably will pose some tough problems and it brings certain dangers in its train. But it must be stressed that all this achievement is the result of the intensive development of data processing in American libraries and information centers.

And thus we come to what might almost be thought of as a small-scale Annus Mirabilis, 1964. To usher in the year, there was the Weinberg Report. Then there was the King Survey, *Automation and the Library of Congress*, followed by the publication of the record of that somewhat unique confrontation of librarians, committed to or actively studying mechanization, and computer experts at the "Conference on Libraries and Automation" held at the Airlie Foundation in Virginia. Another keenly awaited event was the appearance of the *E.J.C. Thesaurus* with its most laudable purpose of standardizing indexing language. However, it may possibly raise more problems than it solves. Finally, there came that August issue of *Index Medicus*, a real breakthrough in automated information retrieval.

A meeting of the F.I.D. in the New World enevitably highlights the question of the place of classification and indexing in the scheme of things. To many, the F.I.D. is identified with hierarchical classification, still trailing clouds of U.D.C. futilities. In contrast, the modern approach offered is coordinate indexing and the *Thesaurus*. Does this represent real progress or are we in danger of losing our perspective in this swing of the pendulum? We seem to be moving towards an elaboration of indexing languages, overlapping and incompatible thesauri tied to one language, somewhat oblivious of the fact that about half the world's scientific literature appears in languages other than English. When Paul Otlet and Henri La Fontaine, at the turn of the century, founded what has become the F.I.D., they saw in classification a great unifying principle, an international language, the standard currency for documentation.

Perhaps Washington II may clarify some of these issues!

London, 9 February, 1965

HERBERT COBLANS, *Editor*
REVUE INTERNATIONALE DE LA DOCUMENTATION

* In an exchange of views, the Editor of AD will submit an editorial for publication in the *Revue International de La Documentation.*

A Note on the Pseudo-Mathematics of Relevance

MORTIMER TAUBE

Documentation Incorporated
Bethesda, Maryland

Recently a number of articles, books, and reports dealing with information systems, i.e., document retrieval systems, have advanced the doctrine that such systems are to be evaluated in terms of the degree or percentage of relevancy they provide.[1–6]

Although there seems to be little agreement on what relevance means, and some doubt that it is quantifiable, there is, nevertheless, a growing agreement that a fixed and formal relationship exists between the *relevance* and the *recall* performance of any system. Thus, we will find in the literature both a frankly subjective notion of relevance as reported by individual users, and equations, curves, and mathematical formulations which presumably provide numerical measures of the recall and relevance characteristics of information systems. This phenomenon of shifting back and forth from an admittedly subjective and non-mathematical term to equations in which the same term is given a mathematical value or a mathematical definition has its ancient parallel in discussions of probability. One cannot, of course, legislate the meaning of a term. It all depends, as Alice pointed out, on "who is master," the user or the term. On the other hand, the use of a single term in the same document to cover two or more distinct meanings, especially when such a usage is designed to secure the acceptance of a doctrine by attributing to it mathematical validity which it does not have, represents a more serious situation than merely careless ambiguity.

One may say, "There is a probability that I will go shopping tomorrow," and in this instance the term "probability" has no mathematical value and refers only to the subjective expectation of the speaker. On the other hand, if it is said that the probability of throwing a head on a coin is ½, our basis is usually the frequency theory of probability, which, as a branch of pure mathematics, has no direct relationship to the psychological strength or weakness of human expectations, especially as such expectations may concern single, non-repeatable events.

This distinction between the strength of a belief and the calculus of probability is made quite clear by Ernest Nagel. He concludes his discussion of this point by noting that:

> In spite of the efforts of De Morgan, Stumpf, and others to assign an interpretation to the numerical value of a "probability" when the latter is understood as strength of belief, no unambiguous criteria have been specified for the equality, addition, or multiplication of probabilities. They are therefore not entitled to employ the calculus of probability upon their interpretation of what probability is because the fundamental operations of the calculus are without any specified content. And finally, this interpretation of probability is completely irrelevant for propositions in applied statistics or physics like "The probability of a male birth in the U.S. is .52," or "The probability of a 10° deflection of an a-ray passing through a film is ¼." As Norman Campbell remarked apropos of the last statement, if anyone proposed to attribute to that probability any value other than that determined by frequency, he would convince us of nothing but his ignorance of physics.[7]

If it is true that the use of any other meaning of probability in physics than the frequency meaning demonstrates an ignorance of physics, then the converse proposition is also true. The application of a metric derived from the frequency theory to a particular case of psychological belief demonstrates only an ignorance of how people use the term *probability* in ordinary discourse.

Although there are a great many recent articles in the literature which deal with the notion of relevance, in the balance of this paper the demonstration of the illegal shift from subjective relevance as a reaction of a user to mathematical relevance as a property of systems will be restricted mainly to the work of the Cranfield Studies and the Arthur D. Little report, "Centralization and Documentation," since these two sources have given the widest currency to the pseudo-mathematics of relevance.

Before examining how the shift takes place, from the nonmeasurable to an absolute measure, it should be established that the shift does occur. A recent report on the Cranfield Studies said that:

> In some ways, the first Cranfield investigation was perhaps less important for the light it shed on four actual indexes than for its contribution

to the development of techniques for the testing of information retrieval systems and for exposing the basic parameters in the operation of an indexing system. By this we mean that the project has produced fairly sophisticated methods for assessing the operational efficiency of existing installations, based on the two crucial factors of recall and relevance. The measurement of the latter is admittedly a very difficult matter, but critics of the Cranfield method have not so far produced an acceptable alternative.[8]

In the first place, it is not required that the critics of the Cranfield method produce a better measure of relevance; their criticism may be directed to the point that relevance as defined by the Cranfield Studies is not measurable. In any case, we have here the admission that finding such a measure is a "very difficult matter." The Cranfield Studies not only recognized the difficulty of measuring relevance, but also indicated that there was a very real question concerning the ability of a librarian or information officer to "recognize a relevant document when he sees one."[9] Having thus established the difficulty of measuring relevance, the Cranfield Studies nevertheless have no hesitation in concluding, "With the aid of the set of documents and the set of questions, it will be possible to test each index language device in turn and so get *precise figures* for their effect on recall and relevance."[10] Some way or another, a vague, hardly recognizable, and admittedly difficult notion has turned out to be precisely measurable.

This type of shift is even more transparent in the Arthur D. Little work. In the report, the discussion of relevance and recall is introduced by the following statement:

> For this purpose, we must first consider the establishment of a crude measure of the relative importance of nearly synonymous index terms to a specified request or "concept." This approach involves somewhat ill-defined considerations, because "relevance" is at best a hazy notion.[11]

But this hazy notion turns up later in the report carried out to the third decimal place. When it is said in this report that the relevance ratio of a search is .036, apparently the reader is expected to forget that .036 is a measurement of "haze."

Having established that the shift does take place in the Cranfield work and the Arthur D. Little report, the next task is to analyze how it takes place. It will be discovered that in discussions of relevance and recall, the axioms of set theory are falsely applied to subjective judgments of relevance, just as in probability discussions the calculus of probability is sometimes erroneously applied to the strength of particular beliefs.

Initially, the Cranfield Studies used the term "relevance" in explaining that "a known relevant document number—that of the source document—was sought in the searching."[12] This usage implies that there is in a particular collection, a single relevant document so characterized in advance of the search. In this case, the relevance relationship is a unique relevance of the document so named to a question or set of index terms put to the system. Certainly, in any test, the number of such previously characterized documents retrieved can be measured absolutely, but such a measure has no relationship to the presumed desires of a searcher for items other than the initially characterized document; nor is there any question of the percentage of relevant documents in the collection since, by definition, there is precisely only one relevant document for each question.

The first change in the notion of relevance occurs in the following passage:

> The figure of percentage retrieval we will refer to as the *recall ratio*. It represents the fraction of those documents which should have been found (being known to be relevant to the question) which actually were found. If R is the number of relevant documents recalled and C is the total number of relevant documents in the collection, then $\frac{R}{C} \times 100$ gives the recall ratio.[12]

Now in this instance, the relevant documents are a percentage of documents in the collection, and it is no longer true that there is only one relevant document for each question. The parenthetical expression in this passage, "(being known to be relevant to the question)," is really the weasel phrase because it is not clear whether the relevance is known in advance or after the fact, and the use of the phrase here seems to tie this definition of relevance with the statement in the previous passage about "a known relevant document number." The subsequent passage, however, indicates that the recall ratio and R itself have changed meaning:

> The recall figures given above should be treated with caution for the following reasons. Theoretically, it is possible for any librarian or information officer to achieve 100% recall (i.e., not overlooking any relevant document) by the extreme expedient of examining every document in the collection (assuming that he will recognize a relevant document when he sees one).[12]

And this change is expressed categorically in the following passage:

> By this time, however, it was becoming obvious that the criterion so far used in the testing of various indexes—namely, the retrieval of a single known relevant document—gave a very incomplete picture of the operating efficiency of an indexing system. Two other factors of extreme importance are: (1) the number of other documents besides the source document relevant to a question which are also retrieved, and (2) the number of irrelevant documents brought out in the course of searching (i.e., noise). It was agreed

that a more significant measure of index efficiency could be achieved on the basis of the two ratios of recall and relevance.[13]

It is clear the discussion now concerns not a known relevant document but any relevant document. This being the case, the expression $\frac{R}{C} \times 100$ does not have a fixed mathematical result because the new meaning of relevance, R, has not been defined. Further, if R has no mathematical value in the expression which gives the recall ratio, then it has no mathematical value in the expression $\frac{R}{L} \times 100$ (where L is the number of documents retrieved in a search) which gives the relevance ratio.

If an attempt is made to determine exactly what is characterized by the term relevance, or what classes or objects are related by the term, it will become very clear that at least two meanings of relevance are being used and that only one of them has been defined. When relevance is used to characterize a document or characterize the relation between a document and the subjective requirements of a searcher, its meaning is as clear as any other psychological predicate; that is, we may accept the supposition that the searcher uses the term relevance with fair consistency in referring to his own responses. When it comes to asserting a consistency of usage for all searchers, we are on much shakier ground. Hence the statement in the Little report that "'relevance' is at best a hazy notion."

In any case, it is clear that in the context of a searcher's judgment, "relevance" is a psychological predicate which describes his acceptance or rejection of a relation between the meaning or content of a document and the meaning or content of a question.

When, on the other hand, discussion shifts to the *relevance-recall* ratio, relevance is assuming to characterize or describe, not psychological states, or the relation of a particular document to a particular question, but systems as a whole. It is supposedly an indexing or I-R system as a total complex which is characterized or ranked according to its relevance and recall performance. It may be said that the system's performance is measured by subjective responses, but on this point two things can be said:

1. A subjective response may be trusted to determine that a certain percentage of items retrieved are not *relevant,* but such a response cannot establish a percentage of recall because the subject does not know what other documents there are in the collection which might be relevant to his question. Hence, if the *recall* performance of a system cannot be a characteristic of, nor derive from, a subject's response, it must be defined without reference to the subject, as a characteristic of the system itself.

   In one experiment,[14] an endeavor was made to assess the recall ratio, by providing to the searcher a random selection of documents from the total store in addition to the documents retrieved by a purposeful match of a question with the store. The randomly selected documents were examined to determine if they included any additional documents relevant to the question. A calculation was then made of the total number of relevant items in the collection in order to establish a recall ratio; however, the sample chosen was so small and the results so ambiguous that the experimenters themselves expressed grave doubts concerning the validity of their results, even though these results allowed a range in the number of relevant items of a factor of) (from 400 to 1,200 relevant documents in a total population of 6,670).
2. It is clear that in both the Cranfield Studies and the Little report, this psychological variation of subjective responses is not regarded as important because of a presumed relevance-recall ratio which presented as an objective property of systems and not an empirical measure of subjective response (especially, as we have noted, no one has ever provided satisfactory empirical measures of such subjective responses.

We come then to the crucial examination of the phenomena of the almost universal acceptance of an undefined ratio and the conclusion in these reports and others that there has finally been achieved a measurable objective property which enables the precise comparative evaluation of systems.

This phenomena, as we have said, arises from a confusion between certain axioms of set theory which all mathematicians accept and all libraries act on, with these demonstrably vague concepts of recall and relevance.

A set b is said to be included in set a when all members of b are also members of a.

The sets—*ab, ac, ad . . . an,* are all included in set a. In terms of the Dewey Classification Systems, with which almost all librarians are familiar, Class 510 Mathematics includes 511 Arithmetic, 512 Algebra, 513 Geometry, 514 Trigonometry, etc.

Suppose a searcher is primarily interested in the Class *ab* but does not know that his subject is also covered in items which are members of Class *ag.* If he asks for *ab,* he will get high relevance and low recall. If, on the other hand, he asks for all *a,* he will increase the number of items recalled because he will get both *ab* and *ag,* but he may effectively decrease the proportion of relevant answers to the total number of items delivered because he will also get *ac, ad, ae, af, ag, ah . . . an.* But note that even in this case, the recall ratio and the relevance ratio cannot be fixed mathematically except by an empirical examination of each case. If *a* had only two included classes, *ab* and *ac,* and both *ab* and *ac* were relevant to

a particular answer, then changing a search from *ab* to *a* would increase both relevance and recall.

In spite of this real and common occurrence, most librarians and information people know that, in general, the larger the set retrieved, the more likely it is that the answer will contain irrelevant material and that the smaller the set retrieved, the more likely it is that relevant material will not be recalled. They also know that because searchers differ and because both the relation of classes and their respective sizes differ, the *art* of searching is to formulate a specific question which isn't too specific, and the art of indexing is to use class designations consistently so that the searcher can have confidence in his *art*.

It is a long and magical step from the abstract properties of sets and the empirical facts of indexing and classification, on one hand, to the claim that a mathematically determinable inverse ratio of recall and relevance characterizes all systems and hence all questions put to all systems.

Hence, the mathematical results of the Cranfield Studies and the Little report must be considered as pseudomathematics, in accordance with Webster's definition of pseudo: "Deceptive resemblance to a (specified) thing —unreal; illusory."

It follows that the conclusions of these reports, namely, the claim of the Cranfield Studies to have discovered a mathematical ratio which will permit the precise evaluation of systems, and the Little conclusion that large coordinate indexing systems must inevitably break down, are alike without substance or merit.

References

1. Don R. Swanson, "The Formulation of the Retrieval Problem," *Natural Language and the Computer*, Paul L. Garvin, ed. (New York: McGraw-Hill, 1963), pp. 255–267.
2. Robert M. Hayes, "Mathematical Models in Information Retrieval," in Garvin's ibid., pp. 268–309.
3. M. E. Maron and J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval," *J. Assoc. for Computing Machinery*, Vol. 7, July 1960, pp. 216–244.
4. C. B. Hensley, T. R. Savage, A. J. Sowarby and A. Resnick, "Selective Dissemination of Information—A New Approach to Effective Communication," *IRE Transactions of the Professional Group on Engineering Management*, Vol. EM–9, No. 2, June 1962.
5. Arthur D. Little, Inc., "Centralization and Documentation," *Final Report to the National Science Foundation*, July 1963.
6. F. W. Lancaster and J. Mills, "Testing Indexes and Index Language Devices: The Aslib-Cranfield Project," *American Documentation*, Vol. 15, January 1964, pp. 4–13.
7. James R. Newton, *The World of Mathematics* (New York: Simon and Schuster, 1956), Vol. 2, p. 1405.
8. F. W. Lancaster and J. Mills, op. cit., p. 9.
9. Ibid., p. 5.
10. Ibid., p. 12.
11. Arthur D. Little, Inc., op. cit., p. 36.
12. F. W. Lancaster and J. Mills, op. cit., p. 5.
13. Ibid., p. 8.
14. C. B. Hensley, T. R. Savage, A. J. Sowarby and A. Resnick, op. cit.

The Aslib-Cranfield Test of the Western Reserve University Indexing System for Metallurgical Literature: A Review of the Final Report*

ALAN M. REES

Assistant Director
Center for Documentation and Communication Research
School of Library Science
Western Reserve University, Cleveland, Ohio

The Aslib-Cranfield Research Project, under the direction of Cyril Cleverdon (Librarian of the College of Aeronautics, Cranfield, England), was initiated late in 1957, and since that time progress has been reported[1] and reviewed[2] on a number of occasions. The development of a testing technique for assessing the performance of indexing systems during the first stage of the Aslib-Cranfield experiments permitted Cleverdon to undertake in 1961, at the request of the National Science Foundation, a test of the Western Reserve University Index to Metallurgical Literature. In the absence of any extensive attempt to measure the performance of an operational indexing system, the results of such a test promised to be of great interest. The Cranfield-WRU cooperative endeavor, as an adjunct to Cranfield I, was viewed as an experiment in the use of an existing test method at the same time as an assessment of the WRU indexing system. The recent publication of a final report on this test by Mrs. Jean Aitchison and Cyril Cleverdon summarizes and analyzes the results.

1. Test Procedures and Results

The design involved the identification of approximately 1,300 journal articles, research reports and patents from the total number of metallurgical documents already abstracted in 1961 at Western Reserve University (WRU) prior to the initiation of the experiment. Documents selected were those published in English and accessible to the Cranfield team. These documents were then indexed at Cranfield, using a modified version of the English Electric Facet Classification for Engineering. All of the facet indexing was performed by Mrs. Jean Aitchison; the WRU indexing was spread over several dozen individuals. Questions based upon documents included within the file were obtained by Cleverdon from a number of metallurgists. Thus, one *source document* existed for each question. Questions were searched by both WRU and Cranfield through their files. Neither the source documents nor other relevant documents were disclosed to the searchers at the time of searching.

If the source document (upon which the question was based) was retrieved, the search was considered a success; 114 questions were searched in this manner at Cranfield and WRU.

In these searches, WRU recalled 94 source documents and failed to retrieve 20, with a success rate of 82.4%. Cranfield recalled 103 source documents in searching and missed 11, resulting in a success rate of 90.3% (p. 12). A sample of the non-source documents retrieved by WRU was taken and the documents submitted to the respective compilers of the questions. Relevance was assessed by the questioner with respect to a scale: Relevance 2—as useful as the source document; Relevance 3—of some relevance; Non-relevant. A finer gradation was tried and abandoned. Of the 295 non-source documents retrieved by WRU and submitted to the question compilers, 13 were assessed as Relevance 2, 47 as Relevance 3, and 239 as Non-relevant.† Extrapolating to the total of 1,703 documents retrieved by WRU in 114 searches, the score becomes:

Source documents	94
Relevance 2	70
Relevance 3	251
Non-relevant	1,288

To compute the total relevance and recall scores, each document in the file was matched against each question by the Cranfield group. For the two indexing systems tested the recall and relevance ratios† were:

* Jean Aitchison and Cyril Cleverdon, *Report on a Test of the Index of Metallurgical Literature of Western Reserve University*, College of Aeronautics, Cranfield, England, October 1963, 270 pp.

† No comparable assessment was made of the Cranfield non-source documents retrieved.

For Relevance 2 and source documents:

	Total Recalled	*Relevant§ Total*	*Relevant§ Recalled*	*Recall Ratio*	*Relevance Ratio*
WRU	1703	156	127	81.4%	7.5%
Cranfield	821	156	133	85.3%	16.2%

For source documents together with Relevance 2 and Relevance 3 documents, the scores are: WRU—recall ratio 75.8% and relevance ratio 17.7%; Cranfield—recall ratio 69.5% and relevance ratio 33.7%.

2. Discussion of Results

An examination of the performance figures compels one to wonder why the WRU system, with its high power of discrimination provided by exhaustive *and* specific indexing, thesaural control and syntactic relationships, did not perform substantially better than the faceted index. If control devices such as role indicators, punctuation linking and thesauri do not materially improve performance, then this has serious implications to the development of information retrieval systems, in that it has been assumed that these devices are desirable if not indispensable to the assurance of high relevance and recall.

The severity of this train of speculation necessitates serious consideration as to what was tested, the nature of the controls surrounding the tests, and design features of the experiments likely to have influenced the results. The interaction between design and results must be critically examined in order to interpret validly the data derived from the experiment. Cleverdon has assumed that his test design was adequate to test his hypotheses and that the integrity of the experiment was maintained in the operational implementation of the test procedures.

3. Test Design

The stated purpose of the test was "the evaluation of the operating efficiency of the (WRU) index, this involving evaluation of the code or index language, and of the intellectual processes of indexing and search programming." There are several variables involved here, yet they were not separated in the experimental design, with the consequence that "indexing" and "index language" were confounded. For example, two of the variables are:

(a) the structure of the index language.
(b) the manner and effectiveness with which the index language is employed.

As to the latter variable, it is evident that the consistency and effectiveness with which each index language is applied are of some significance since it is possible that the *worst* index language utilized by the *best* indexer may tend to approximate or even excel in performance the *best* index language utilized by the *worst* indexer. With respect to the former variable, measures of performance relating to the structure of index languages (effectiveness of devices such as role indicators, interfixing and the like) need to be performed with the indexing process (variable (b)) held constant between languages. This necessitates the input of a "common body of concepts" into each of the index languages tested. In tests of index languages, all other variables must be held constant except for the index language which is the variable to be tested.

As the result of the Cleverdon experiments, we now know a great deal more about the complex interaction of variables within and between indexing systems. In the Cranfield-WRU test, indexing was performed with no controls with respect either to depth of indexing or to the concepts extracted from documents and translated into the two index languages. Therefore, it is difficult to determine the degree to which the index languages were compared as distinct from the indexer's selection. Moreover, the Cranfield indexing was performed by one person on an experimental basis, whereas at least two dozen persons were involved in the indexing of the test documents at Western Reserve over a considerable time period.

This distinction, now apparent, between indexing as a process and the index language utilized has subsequently proved most useful to us at Western Reserve in our Comparative Systems Laboratory. In the Cleverdon test, however, we do not know the extent to which recall and relevance varied as a function of indexer selection or as a function of index language, since the selection was not held constant between the two systems; moreover, we do not know the extent of variation within one indexing system with respect to differences in background, skill, academic accomplishment of indexers. Was the complex WRU index language poorly applied and manipulated; or is there evidence for structural weakness?[3] The test design does not permit an answer and leads to much confusion.

In the same manner that a lack of control was applied to the indexing, it is now apparent that all other variables were not held constant. Assuming that an information retrieval system is "an integrated assembly of components that interact cooperatively to perform a predetermined function for a specific purpose,"[4] it follows that tests of one component of the system, such as indexing, necessitate the neutralizing of the effect of all other component elements. Question analysis in particular was not held constant between the two indexing systems.

The role of searching in these experiments is great, considering that 67.1% of the WRU recall failures were due to the search programs (p. 20) and that 40.2% of the WRU non-relevant documents were retrieved due to the same reason (p. 27). Cleverdon states:

> Cranfield programs had better concept matching than WRU, omitting fewer concepts, adding

† Recall ratio equals $\frac{100R}{C}$ where C equals the total number of documents in the collection which have an agreed standard of relevance to a given question, while R equals the number of documents retrieved in a single search. Relevance ratio equals $\frac{100R}{L}$ where L equals the total number of documents retrieved in a single search.

§ Assessed as *a priori* relevant by Cranfield.

> fewer concepts, whether more or less specific than the question or implied in the question . . . the effect of using Cranfield search programs against WRU indexing for all cases where WRU failed to recall relevant documents shows that Cranfield programs would have retrieved 71 of the documents not found by WRU.

Failure by WRU to match question concepts with program concepts resulted in not only loss of recall but also in the retrieval of large numbers of non-relevant documents. Use of the Cranfield programs would have achieved the exclusion of 33 out of 50 non-relevant documents analyzed by Cleverdon (p. 39).

Matching was indeed possible, and the WRU searches could have closely followed the words contained within the question. Instead an attempt at question analysis was made in some, but not all, instances in order to arrive at an appropriate generic-specific level of concepts and to determine concepts validly related either collaterally or synonymously with the question concepts. In normal operational procedures, these decisions are made by recourse to the questioner in order to bring a formalized representation of the question into closer coincidence with the information need which motivates a question. Without such recourse, question analysis was based upon guessing, and the WRU staff played the dual role of questioner and searcher.

The significance of this is not so much to explain the failures but rather to point out that question analysis was not held constant in the tests, with the result that different prescriptions were searched against the two files. It is now apparent that three alternate procedures exist:

(a) Program words in question (i.e., no question analysis).
(b) Question analysis by searcher (with no recourse to the questioner).
(c) Question analysis with recourse to the questioner.

Finally, the so-called *synthetic* questions used in the experiment precluded negotiation with the questioner and served to test only the ability of the indexing systems to retrieve documents which matched textual statements or concepts selected from a document and inverted to form a question. Accepting a question at its face value tests the capacity of a system to match the wording of a question and not the information need. Tests must therefore be designed with live users, *natural* questions and question analysis, to permit the matching of output with the underlying motivational need.

4. Significance of the Experiment

The conclusions drawn by Cleverdon from the Aslib-Cranfield test on the WRU Index seem to be (to paraphrase Cleverdon):

> The trading relationship between relevance and recall hypothesized does in fact exist; interfixing is preferable to role indicators and punctuation levels in that the latter did not perform very effectively; the WRU indexing was too exhaustive; the considerable effort put into the semantic code seems hardly justified either operationally or economically; there is no evidence that a computer is more effective as a retrieval mechanism than a conventional card catalog; the very devices used to ensure high recall and relevance militated in many instances against the successful performance of the system; there is little of an ersatz quality in synthetic questions (notwithstanding Fairthorne and Rees); the validity of the test method is proven.

Without arguing *in extenso* each point made by Cleverdon, it is evident that the Cranfield test method needs further investigation with respect to its underlying design. In its application to the WRU indexing system, it is not clear as to what was tested. The complexity of the indexing process and use of index languages within the framework of a total information retrieval system is such that it is difficult to be confident that the mass of resulting data really relate to the hypotheses tested. Are any of Cleverdon's experimental results reproducible either at WRU or Cranfield with different indexers analyzing the same body of documents, with yet other persons searching the same questions against both the original and new files? Can the same experiments be performed by persons other than the Cranfield or WRU personnel to achieve identical results with the performance figures in Cleverdon's report? One doubts whether the design provided for the rigidly controlled conditions which ensure reproducibility of results. The lack of such control leaves lingering doubts as to the attributability of the data to the original hypotheses (which themselves were never sufficiently formulated).

What has the Cranfield work shown? In Cleverdon's opinion, the test proved the validity of the test method:

> It is considered that this test has shown the possibility of ascertaining, within reasonably precise limits, the operating efficiency of an indexing system . . . [and] that the analysis which can be done is sufficient to show why the system is operating at a particular level and thereby to show how it can be operated more efficiently or economically.

This reviewer holds that the last word has not been spoken by Cleverdon on testing. The great value of Cleverdon's contribution lies in the area of test methodology rather than in the experimental results. The Cranfield work has exposed in great detail the problems involved. Awareness of these paves the way for the formulation of further hypotheses which more closely approximate reality and for the more effective design of experiments to test these hypotheses with rigid controls.

Only in this manner can we know what to test and how to design tests with an awareness of what in fact is measured.

With all imperfections taken into account, nonetheless Cleverdon has with these tests made a significant contribution to our knowledge of retrieval systems and has pointed the way toward future research.

References

1. C. W. Cleverdon, *Aslib-Cranfield Research Project: Report on the First Stage of an Investigation into the Comparative Efficiency of Indexing Systems* (Cranfield, College of Aeronautics, September 1960); *Interim Report on the Test Program of an Investigation into the Comparative Efficiency of Indexing Systems* (Cranfield, November 1960); *Report on the Testing and Analysis of an Investigation into the Efficiency of Indexing Systems* (Cranfield, October 1962).
   See also C. W. Cleverdon and J. Mills, "The Testing of Index Language Devices," *Aslib Proceedings,* Vol. 15 (April, 1963), pp. 106–130.
2. The first stage of the Aslib-Cranfield Project was reviewed by John O'Connor in *Journal of Documentation,* Vol. 17 (December 1961), p. 261; and by Phyllis Richmond in *American Documentation,* Vol. 14 (October 1963), pp. 307–311.
3. The reasons why role indicators, punctuation linking and so on did not perform as well as was anticipated are given in:
   Alan M. Rees, "Review of a Report of the Aslib-Cranfield Test of the Index of Metallurgical Literature of Western Reserve University," Center for Documentation and Communication Research, WRU, October 1963. (Lecture given at Aslib, London, October 16, 1963.)
4. William Goffman and Vaun A. Newill, M.D., "The Complex Problem of Medical Information Retrieval: I. A Methodology for a Comparative Systems Laboratory," January 1964, p. 3 (to be published).

$\mathrm{F}^1\mathrm{a}^1\mathrm{c}^1\mathrm{e}^2\ \mathrm{I}^2\mathrm{t}^3\mathrm{s}^1\ \mathrm{N}^3\mathrm{o}^1\mathrm{r}^2\mathrm{m}^1$

RALPH R. SHAW

Dean of Library Activities
University of Hawaii
Honolulu

The title of this paper, as indicated above, is *Information Centers*. While it appears to read *Face Its Norm* with numerical superscripts on all of the letters, it should not be too difficult to write a computer program which would recognize the superscripts as indications of the frequency with which each letter is to be written, and then, by using all permutations and combinations of these letters, in the frequency indicated, to come up with *Information Centers*, in a relatively short time, as one of the possible choices. The program could, of course, be refined so that we would not actually have to write the superscript numeral ^{1}, since we could instruct the machine to consider a blank space between two letters as meaning the numeral 1. I should like to point out that this new game, or what I believe to be a new game, not only has the advantage of being a fresher and more scientific-looking version of Scrabble, but it took only about a half-hour's time to work it out. (The programming for doing the rest of it is so obvious and so easy that it could be done by the computer staff that is there anyhow, so it would not cost anything and is not included). I must admit that it takes more typewriter strokes to write the title in this new form, but I submit that this technique gives me two titles (the 24 strokes* are only 5 strokes more than would have been required to play it straight, not counting the computer time because it is sitting there anyhow) and they both make sense. The problem is, what sense do they make? And that is the subject of this discourse.†

The characteristics of information centers, as they appear from reviewing the published and unpublished literature and from personal observation, are the following:

1. They deal in science information.
2. The current concept limits them to narrow areas of specialization.
3. Each is closely associated with a group of working scientists in its area of specialization.
4. While they collect and organize and index artifacts that contain data, the common thread in all attempts to define their functions and to differentiate them from other agencies that perform some or all of the same functions, is that their prime function is the application of a high order of judgment by scientists who have the highest possible substantive competence in the particular area of specialization to organize, summarize, abstract, evaluate or otherwise manipulate this raw material, retrieving from it the essence of information pertinent to the particular problem at hand that is contained in a large corpus of raw informational materials, and to communicate this essential knowledge to their peers.
5. They are designed and operated by scientists for scientists.
6. They use (or allegedly use, or might sometimes use) some form of hardware.

While the characteristics enumerated are of varying degrees of importance, and the stated objective may vary widely from the norm of actual operations or achievement, there can really be no difference of opinion about the need for improving information services to science, technology, and to all other areas of scholarship as well. Furthermore, whether we agree in whole, in part, or not at all, on any particular device selected for achievement of this goal, we must respect and admire the imaginative and courageous efforts that are being made to do something about this important social need.

There is nothing fundamentally new about the concept of information centers as defined by any, or any combinations, of the criteria listed above.

This type of activity has been carried out over many years in special libraries, industrial research complexes, and even in some of the larger public libraries. It has been quite common, in libraries such as that of the Department of Agriculture, for the bibliographical staff to do the preliminary search to gather material that may be of pertinence to the topic under study, basing this search

* There will, no doubt, be those who will want to charge us with the extra time for the eleven platen shifts, but that is one of the penalties for trying to do creative work; it is always subject to carping criticism.

† It has required only a little more than 300 words to explain this device for saving 8 typewriter strokes, and that is probably better than par for the course.

on specifications developed with the cooperation of the subject specialists in the Department, and then calling on the subject specialists for final evaluation of the materials contained in the literature and for the preparation of a summary of the state of the art.

The fact that the concept is not new does not, however, suggest that it is not important, and the fact that it has been done sporadically in the past does not in any way indicate that the greatly accelerated program now developing may not, simply by more effective and extensive prosecution of this approach, achieve new orders of intellectual support of scholarship. It will do that, however, only if it avoids the mistakes of the past and if the program is designed and executed so that it does in fact, rather than simply in slogan, make a contribution to science. To outline the steps that appear to be necessary to make information centers achieve their promise (which will not be achieved simply by pouring money and machines into them), it might be well to review the basic problem and the special characteristics attributed to them so as to arrive at an understanding of their potentials and the route or routes by which these might be achieved.

There have been many attempts to differentiate documentation centers or information centers from libraries. These have not succeeded since most of the differentiations have been based on careful selection of concepts to fit the case that the definers were trying to make. Even if we were to define documentation as a specialty within librarianship, however, there would still be an additional facet of the work performed in information centers as defined above, which, while it could be done *in connection with a library*, could not be done by one who would be doing it *as a librarian*, no matter how broadly we define librarianship. Thus, there is a basis for differentiation.

Perhaps the best way to approach this is by comparing these functions with the task of dealing with raw intelligence as compared with the task of producing intelligence. In this context, we are using the term intelligence with a lower case *i* (which means sans false whiskers). Raw intelligence happens when data, regardless of its physical form, is collected, roughly organized, and made available to trained minds. Intelligence is created, hopefully, when these trained minds can reorganize, sift, compare, the raw intelligence with what is known in the state of the art, and can make magic leaps of genius—thousands of daily little leaps and some great ones—that create relatively reliable new intellectual content, i.e., intelligence.

The collection and preservation of raw intelligence may be done purposefully or it may be accidental. For example, Columbia University has a series of bound books in which the library kept records of loans as they occurred more than 100 years ago. Analysis of this record could be made to create intelligence about use of the library by scholars a century ago. Similarly, the cash books kept by colonial farmers are raw intelligence. Their utilization by historians and social scientists can result in provision of socially useful, new knowledge about the history, social structure, and cultural anthropology of our country.

Approached from this point of view, the raw intelligence function and the creative intelligence function are closely interwoven and have an almost infinite range of possible interrelationships. It is not feasible, therefore, to decide on the method before deciding the purpose. It is quite possible that in deciding on a very narrow purpose, which may be required if we are going to have materials in close relationship with those who are capable of creating intelligence from the raw materials, we may lose more than we gain because we do not know much about the input into a scholar, which makes it possible for him to create intelligence, or whether it dare be limited to only those items specifically assigned to him for judgment at any given moment of time. Such studies as we have had of the reading habits of scientists indicate that probably half of their reading is completely outside of their field of specialization or clearly related fields, and it may very well be that it is an exceedingly important half in terms of equipping them for sounder judgments in their own field of specialization. This needs more study.

Detailed examination of the characterizing features of information centers indicates that the crux of their identity lies in their large scale use of people of deep substantive competence, each in his own field of specialization, to create and to disseminate intelligence from a collection of raw intelligence materials, regardless of whether this is done in a distinct administrative entity or as a means for achieving the desired result without considering the formal administrative structure.

Thus, going back to our list of distinguishing characteristics, the first one, which limits information centers to the field of science, is not an indispensable requirement. They may deal in science or they may deal in linguistics, psychology, or other areas which are of importance to society and to the end product of science. It is not profitable, however, to argue this one. Let us agree that they may deal with science, among other things.

The current concept, as noted in the second criterion, limits them to narrow areas. Since no one can be a specialist in broad areas, this would appear to be an obvious requirement, but it is an obvious requirement only insofar as the quality of judgment is concerned and may be hampered by making the range of the collections narrow. In some areas we must draw from many disciplines in order to gather the materials that are then subjected to critical evaluation and judgment in solving a particular problem. Thus, here, as in the case of departmental libraries on the university campus, there is convenience in having the material on a specialized field together, but there is danger that much pertinent material may be missed because it can not be or is not duplicated in the

specialized collection. Furthermore, we do not know how the creative mind works and we do know that the studies of reading by scientists and technologists have all indicated that a very large part of the normal reading of scientists is outside their specialties or related specialties. Thus, while the person applying judgment to raw intelligence materials must have access to these materials, there is some danger in defining the range of material to which he has access too narrowly.

Criterion three, the need to have the information center associated with a group of working scientists, is obviously an important criterion and is one that is met, at least in part, by every special library. Again the intellectual result is the important thing rather than the physical object or form, and it may be that this could, at least in some cases, be better achieved by bringing a subject specialist into the National Library of Agriculture or the National Library of Medicine, and providing him with legmen to collect the material for him from these libraries and others, rather than by hoping to have the best people in all the sub-disciplines that are to be covered working in a particular locus surrounded, in each case, with all the materials that might be pertinent as raw intelligence.

There would appear to be little question that the fourth criterion states the essence of the concept of information centers: the application of adequate scientific judgment to raw intelligence to create intelligence.

Requirement number five appears to be moot. It requires that information centers be designed by and operated by scientists. While the judgment portion of the operation unquestionably requires scientists, the administrative function calls for a high order of administrative ability applied to other than scientific operations and if scientific, to many operations outside the field of competence of any single scientist. Thus, while the administrator of a science information center may have been a scientist and may have substantive competence in a very narrow portion of the functions performed by the center, that does not, per se, qualify him to administer the center. With all due respect to science and scientists, a man may be a great scientist without knowing or caring anything about ordering books or documents or data sources of other types, or organizing them for use, or supervising staff, or any of the other myriad skills that are involved in administering an information center. This argument about whether we should have subject specialists or literature specialists trained in the art of administering information resources has been going on for half a century. It is not a new argument. No objective data have been supplied to support either side of the argument in the thirty years since one of my students at Columbia made up a bibliography of a couple of score articles on the subject. One of the better studies in this area is that done by Richard Dougherty, in his doctoral dissertation at Rutgers, which dealt with "The Scope and Operating Efficiency of Information Centers as Illustrated by the Chemical-Biological Coordination Center." This covers argument number six (i.e., the implication of mechanization) as well as number five, the administrative problem. Here we have a case in which the decision to base the information center on a particular mechanical approach and the failure to develop a structure that permitted administrative freedom in planning and operation of the center doomed it despite the fact that it had a fine substantive staff and a head who was a scientist. The important requirement is that we have staff competent to do each of the kinds of jobs that needs to be done and that we use the tools and systems that are most suitable, rather than predetermining the kinds of staff and tools to be used based on labels or slogans.

This calls for flow-process charting the work done and to be done, determination of the kinds of skills required for each part of the task, developing the best possible system and determining the best tools available for making sure that the system is in fact carried out effectively. Once we have done that, determination of the administrative structure to be used becomes relatively simple, and it may differ from case to case. There may well be some types of information center operations that are best carried out by building up a separate collection at an isolated point, and others may well be achieved with greater effectiveness by temporary assignment of teams of scientists or other scholars to a location in the Library of Congress or some other great collection, with supporting staff located there to relieve them of the work, other than judgment, that needs to be done. This might well combine skills in the vast materials handling job and in the important intellectual work that needs to be done, without wasting scientists on levels of work for which they are not trained and without using information specialists or librarians or documentalists for making judgments for which they are not qualified.

One more *caveat*. In evaluating programs of this type, it is necessary to bear in mind that we are concerned with two types of efficiency. The paramount consideration is program efficiency. If the program needs intellectual support and must have it to achieve its purposes and if there is only one exotic way in which this can be done, then it must be done that way and questions of cost become secondary. In most cases, however, we have choices of methods, systems, tools and staff, and in these cases, within the framework of the mission to be achieved, it is possible and necessary to give proper attention to operational efficiency. This latter is by far the more common case, and assuming that the same end result is achievable by two or more methods, the burden of proof should, in common sense, be on the person who wants to do the job the more complicated or more expensive way.

And now, returning to our title, let us indeed face up to the norm and examine what is, in fact, done in information centers, rather than what we talk about when we attend meetings such as this one.

There have not been many objective studies of what actually is done in information centers and those that have been published are open to argumentation and debate.

There is generally a great gap between claims and achievement, or between the exceptional cases cited as the norm and the norm. It is doubtful that the average scientist or engineer really is greatly concerned about the great mass of literature, and many appear to be quite content to get along with cookbooks (i.e. handbooks that tell them what formula to use). There are frequently great divergences between the alleged performance and the actual performance in the average case in this and in other fields—and this is quite understandable.

A medical research library like the NIH Library obviously needs people who can understand the meanings of the words in a biomedical document or text, and this quickly gets translated into an alleged requirement that all the library staff must have medical, or at least science, background. There is unquestionably need for some people with science background, but that hardly goes all the way down to the janitor. In fact, my examination of the reference questions actually handled by the reference staff of the NIH a few years ago showed that about 90% of them were questions that could have been (and largely were) handled by general reference staff from general reference tools, with the staff requiring no more science background than we would find in a typical graduate of a liberal arts college whose major was in a field other than science. This indicates that the either-or argument on staff requirements may well be naive and misleading and may actually delay the development of useful information centers.

Similarly, a few years back, when Battelle was operating its Titanium Information Center, we had the claim that physicists, chemists, etc., had to do the work. Conversations with people working on the project as well as examination of reports of the project, indicated that we had physicists and chemists performing order routines, cataloging, answering routine reference questions, and doing a number of other necessary tasks for which they were not trained or particularly well qualified, and which could have been done by people trained for these jobs without waste of scarce scientific manpower. Of course, the later evolution of the Battelle program has largely corrected this situation, but do we really have to have ontogeny recapitulating philogeny time after time and in case after case, or will we ever grow up to modern management—including, especially, the design of such operations so that we may build them on facts rather than on slogans, and particularly to the level of administration at which we identify the actual work to be done in each task and have it done by personnel at the lowest level at which it can be performed, including provision of the higher skills where they really are needed without wasting them where they are not?

Let us take just two more examples by way of driving this point home. At Agriculture, for years we had a wonderful Russian bibliographer who spent a large part of her time writing out the roman alphabet transliteration of Russian titles for the typist to copy. One of our typists suggested that she might be able to help by transliterating directly and a great crisis developed. When the dust settled it was easy enough to establish that the transliteration equivalents could be fitted onto a 3 x 5 inch card and that one did not really need a Ph.D. in Slavic languages (or even proficiency in machine translation) to transliterate. Result—one typist, at the going rate of pay, looking at cyrillic titles and typing the English transliteration faster than she could decipher the distinguished handwriting of our distinguished bibliographer, and one bibliographer happily doing a good deal more of the high order of intellectual work that only she, of all the people on our staff, could do.

As a final example, let us take the Cohen and Craven study of Science Information Personnel, sponsored by the Modern Language Association. On pages 5-6 of this report you will find a table and a chart of the "elements of science information work" they found to be performed in 207 federal, industrial and research agency information centers that they studied. They tabulated these in terms of frequency with which these elements are found in the 207 agencies, and the only elements found in 25% or more of the information centers, in the frequency with which they were reported, are the following: (1) literature searching, (2) reference work, (3) subject analyzing, (4) abstracting, (5) administering, (6) descriptive cataloging, (7) translating, and (8) selecting materials. There are none of these elements that are not common to special libraries. To be sure, frequency must not be confused with importance of the functions; and other functions may well be more important than any of these. It is an inescapable conclusion, however, that a very large percentage of the work common to information centers is not dependent upon a high order of substantive competence in narrow specialties in science or technology, and that training in science has not been established as the primary attribute required for effective performance of many important parts of the work of the information center.

If we can substitute fact for slogan, and if we can outgrow an either-or approach to assignment of tasks, we should be able to develop information centers that will blend the necessary substantive competence with bibliographical, indexing and other skills, and the necessary administrative and management skills, in suitable development or selection of systems, methods and tools, so as to make the information center a powerful and efficient tool for intellectual support of scholarship in science and technology, and in other fields as well.

The speed with which we approach this goal depends on you.

Improvement of the Selectivity of Citation Indexes to Science Literature Through Inclusion of Citation Relationship Indicators

Citation indexes to large bodies of science literature can often list far more citing references under the known cited reference than the user can afford the time to look up. By providing some additional information, beyond the minimal association of the citing reference with the cited reference, a citation index could provide the means for the user to select from a long list those citing references which are most relevant to his immediate search requirement. Means of providing this selectivity are discussed. Particular attention is given to the possibility of adding short codes to the citation entries which would be informative on the way in which the citing publication is operationally related to the cited one (this method is an integral feature of the Shepard's legal citations). A scheme of citation relationships of potential value to users of science literature is presented. These relationships were tested on a sampling of physics literature. The suggested categories include indicators of the relation of the citing reference to the scientific process in general, as well as indicators of its relationship to the cited reference in particular. Assignment of the categories to a citation requires the exercise of judgment, as in subject indexing, but does not involve the use of subject terminology. An illustration is provided of the application of citation relationship indicators to an excerpt from a citation index to physics literature.

BEN-AMI LIPETZ

Carlisle, Massachusetts

This paper is the result of an exploratory study of the possibility of providing for greater selectivity in citation indexes to science literature to permit the index user to distinguish between the citing references which are most pertinent and least pertinent to his search problem, without having to look up all of the citing references in a library. This study was carried out under the sponsorship of the American Institute of Physics Documentation Research Project, with support from the National Science Foundation. It was an outgrowth of concurrent work on evaluation of the usefulness of citation indexing in physics.[1]

• Citation Indexes and Selectivity

In the various experimental and commercial citation indexes to science literature which have been compiled in the last five years, emphasis has generally been placed on reducing the compilation process to a completely routine and mechanical procedure in order to make the fullest possible use of automatic data processing equipment and computers. Human participation in the process has been restricted to the act of keypunching the citing reference and the cited reference from the source literature, to create a punched card or paper tape record. The accumulated machine readable records are sorted by automatic means into a new arrangement by the cited reference (e.g., by author, or by journal, volume, and page) and are printed out automatically for publication. By this process, it is possible to create huge citation indexes rapidly and at a low cost, in terms of the number of citations indexed per dollar of compilation cost.

From the viewpoint of the scientists and engineers who are expected to use the citation indexes, however, there are some distinct drawbacks to cost cutting in citation index preparation through extreme emphasis on mechanization. One of the drawbacks[2] is that a large citation index can frequently frustrate the would-be user by giving him far more references to look up in a library than are warranted by the time available or by the importance of the search question. This happens when the user looks up the starting reference which he knows to be pertinent to his interests and finds that it has been cited by dozens or hundreds (or, in the foreseeable future, thousands) of subsequent publications. While not all references are cited very heavily, enough of them are to make it a serious problem, and one which will get worse as citation indexing in science continues. The dilemma of the user of the citation index is that he knows from experience that only a fraction of the references which cite his starting reference will be relevant to his search requirement;

yet he cannot find out which references are relevant and which irrelevant without conducting a time-consuming, hit-or-miss library search.

If a citation index were to provide the user with not just all the references citing the known reference, but with some means for distinguishing at once between the citing references which are relevant or irrelevant to his immediate interest, he would be able to cope with the plethora of irrelevant (to him) citations, and go directly to the relevant material in the library. The means afforded for making such a selection in the citation index produced by highly automated procedures are very limited. With some types of citation index format the user may be able to do some elimination on the basis of the year of publication if there is a logical cut-off date in his search question. He may also, with some types of format, be able to eliminate some citing references on the basis of their authors' names or their places of publication, if he is fortunate enough to be already highly familiar with the literature of the research area in which he is interested. These latter devices would not work for the person just changing fields or the novice, just the people who have potentially the most to gain from the use of citation indexes. With very long lists of citing references, even experienced specialists would sometimes be hard put to eliminate enough *noise* from the lists to warrant looking up all the remaining references.

To introduce the means for higher selectivity in a citation index requires that the bare citing reference be supplemented with additional information which, in one way or another, reflects the content of the citing reference as it relates to typical search problems. In other words, each entry in the citation index must be expanded to include one or more additional data categories. Obviously, compilation costs, printing costs, and index bulk must also increase as a consequence. Thus, careful study is necessary in order to be sure that the advantages of any given change in citation index content outweigh the unavoidable disadvantages of the change.

• Approaches to Improved Selectivity

Three approaches are available for expanding the information content in a citation index entry. Each has definite advantages and disadvantages.

One approach is to copy into each index entry some portion of the citing publication which will indicate the content of the citing publication. The obvious portion to copy is the title. This keeps the indexing process strictly mechanical and holds down the input processing costs; however, it greatly increases the length of each entry and, therefore, greatly increases the index publication cost and the physical bulk of an index. From the index user's viewpoint, one can question whether titles are usually sufficiently informative to permit identification of the relevant references in a long list of citations. Including an author abstract of each citing reference would provide a better basis for selection but would vastly increase the index publication expense.

A second approach is to perform subject indexing or key-term indexing of some sort on each citing paper and to include the subject terms or descriptors with each citation index entry. Such indexing could presumably be carried to any depth necessary to give adequate selectivity to the citation index. Like the first approach, it can greatly lengthen each index entry and thus increase the size and publication cost of the index. Furthermore, by introducing a procedure requiring the exercise of judgment by skilled personnel, it greatly increases the initial preparation cost of an index. It should be emphasized that all of these disadvantages may be tolerable if the result is an index which provides sufficient advantages to the users. On the other hand, one may properly ask what advantage a citation index has over the currently available conventional subject indexes in science if it requires incorporation of a conventional subject approach to make the citation index usable.

The third approach is considered in the remainder of this paper. It does not attempt to provide a description of the citing article *per se*, but rather to describe the precise relationship of the citing reference to the cited reference. This approach avoids any explicit statement of subject content. Instead, it attempts to codify the important ways in which an existing paper may be acted upon subsequently, and to assign the appropriate codes to a citing reference to indicate its precise continuity relationship with the cited reference. It bears a similarity to subject indexing in that skilled human judgment is required to determine the citing relationship. On the other hand, it lengthens a citation index entry by merely a few code symbols and does not add as much to publication costs as the first two approaches given above.

Although the concept of citation relationship indicators may be new to people working in science literature, it is very familiar in the legal profession, the home of the citation index. Relationship indicators are an integral feature in the various Shepard's citation indexes of judicial decisions, the generations-old legal series whose phenomenal success inspired the current interest in citation indexes for science. In Shepard's, one may look up a known court case to find out whether and where it was referred to in the decisions in subsequent cases. But, to help the lawyer avoid unnecessary look-up of references which are not relevant to his requirements, Shepard's adds appropriate code symbols to the reference of the citing decision, to indicate whether it applied the earlier decision, or whether it affirmed it, or clarified it, or modified it, or overruled it, etc. This makes the citation index a vastly more powerful searching tool than it would be without such indicators, yet only slightly more bulky.

Relationship indicators have so far been omitted from science citation indexes mainly because they require the use of skilled and costly human labor. In science docu-

mentation work there tends to be an obsessive drive to reduce all operations to the simplest, unthinking data processing routines. This is understandable if it leads to real over-all economies, but not if it only reduces costs for the information producer by imposing high costs for the information user at the other end of the communication chain. A potentially valuable feature of science citation indexes should not be ignored merely because it happens to require the exercise of brainpower and cannot be automated easily. Therefore, it was decided to take a closer look at citation relationship indicators to clarify how they might be applied to science literature to help the research scientist and the research librarian.

• Relationship Indicators as Applied to Science Citations

Just as there are no absolute and exclusive terms by which a document may be fully described in subject indexing, there are no absolute and exclusive categories which fully describe the relationship of a citing publication to the cited publication. Relationship categories, like subject descriptions, may be multiplied, modified, qualified, or subdivided to infinity. Decisions on how far to go in relationship coding, as in subject indexing, must ultimately be determined by economic factors which force the conscious or unconscious matching of indexing depth and scope to the practical value of the results. Because lawyers and scientists have different professional problems and different working practices, there is no reason to assume that all of the relationship indicators used by Shepard's can or should be applied to science citations; nor should it be assumed that the relationships found in Shepard's are the only ones worth considering for science literature. Therefore, in attempting to draw up a list of citation relationships which are meaningful and potentially useful in scientific research, Shepard's was used only as a suggestive guide. It was supplemented by ideas drawn from previous work by the present author on the nature of the scientific process.

The feasibility of analyzing relationships between citing and cited publications into a fairly small number of categories of probable value to physicists was investigated through inspection of actual papers and their bibliographies in physics journals. Relationships were analyzed and recorded for about 750 different citations in about 60 different papers in recent volumes of *The Physical Review* (Vol. 128) and *Soviet Physics—JETP* (Vol. 15). Spot checks were made in other physics journals. The analysis was based on a preconceived scheme of useful relationships which had to be modified somewhat on the basis of actual trial. The descriptive scheme which evolved now appears to fit virtually all circumstances, at least in physics literature, but should still be regarded as experimental and subject to further change.

The categories of relationship which are being tested are specifically intended for use in large science citation indexes to help the user select from long lists of citing references the ones which are of most probable pertinence to his immediate requirements. To be most economical, the categories of relationship should be susceptible to designation by short codes. To be reliable, they must be more or less equally recognizable to different literature indexers (who would presumably be trained in a manner analogous to subject indexers). And, above all, to be of any value at all, the categories must be meaningful in scientific work.

At present, 29 categories are being used. They can be thought of as belonging to four major groups (Table 1):

TABLE 1. Citation Relationships in Science Literature

GROUP ONE. Original Scientific Contribution or Intent of Citing Paper
1. Description of observed phenomena
2. Data transformation
3. Explanation
4. Hypothesis or theory
5. Calculation from theory
6. Prediction
7. Definition or notation
8. Statement of experimental technique

GROUP TWO. Contribution of Citing Paper Other than Original Scientific Contribution
9. Review article
10. Bibliography
11. Data cumulation

GROUP THREE. Identity or Continuity Relationship of Citing Paper to Cited Paper
12. One or more authors in common
13. Same text
14. Abstract or condensation
15. Erratum
16. Continuation
17. Precursor
18. Inclusion

GROUP FOUR. Disposition of the Scientific Contribution of the Cited Paper in the Citing Paper
19. Noted only
20. Distinguished
21. Reviewed or compared
22. Applied
23. Improved or modified
24. Replaced
25. Changed the precision (plus or minus)
26. Changed the scope of applicability (plus or minus)
27. Questioned
28. Affirmed
29. Refuted

GROUP ONE is a departure from the precedent in Shepard's in that the categories do not indicate direct relationships between the citing paper and the cited paper. Instead, they describe the citing article's *scientific contribution or intent;* that is, its operational relation to science in general. Of the eight categories in this group, six are the "basic" elements of scientific achievement postulated previously.[4] These are: description of observed phenomena; explanation; hypothesis or theory; prediction; definition or notation; statement of experimental technique. To these have been added two more categories

CITATIONS OF PHYSICAL REVIEW

VOL. 73 (1948)

16 1 TP KOHMAN
62SET14 1265 3 2 KA PETRZHAK

30 1 B HOFFMANN
57SET 5 642 19 1 A POPOVICI

46 3 L APKER
56SET 3 159 17 2 ND MORGULIS
58STP 3 1752 4 2 PG BORZIAK

51 1 GC WICK
57SET 5 323 6 2 DF ZARETSKII

83 2 L GOLDSTEIN
61STP 6 38 4 4 VE GOLANT

111 4 DJ HUGHES
56SET 2 301 3 3 LE LAZAREVA

127 3 HW LEWIS
61SET12 527 5 2 VM MAKSIMENKO
61SET12 554 1 2 BT VAVILOV

155 1 C KITTEL
56SET 2 703 6 1 AI PIL'SHCHIKOV
58SET 6 1155 12 1 VL GUREVICH
60SSS 2 332 4 1 SV TYABLIKOV
61SET12 1111 1 2 AS SHUR
61SSS 2 1805 12 1 SV TYABLIKOV
62SSS 4 1165 8 1 EN YAKOVLEV
62STP 7 316 1 2 DI MASH
63SSS 4 1282 5 1 VL BESHIDZE
63STP 7 820 6 2 DI MASH

162 2 C KITTEL
55SET 1 37 8 1 SA AL'TSHULER
61SSS 2 2009 2 1 AM VASIL'EV
62SSS 4 175 3 3 VK KONYUKHOV

241 4 NP HEYDENBURG
62SET15 243 1 4 IB TEPLOV
63SET16 47 2 1 AM PETROV

256 4 CT LANE
57SET 5 819 5 1 VM KUZNETSOV

257 2 EP HINCKS
63SET16 1073 1 1 B PONTECORVO

262 2 E MCMILLAN
55SET 1 576 1 2 PS BARANOV
56SET 2 677 13 3 VI GOL'DANSKII

266 1 PV HOUGH
59SET 9 1263 17 2 AA VARFOLOMEEV

268 2 BT MATTHIAS
60SC 4 54 11 2 EG FESENKO

297 1 H MARGENAU
55SET 1 205 3 1 VM FAIN
56SET 3 14 8 1 GM PATEIUK
56SET 3 147 7 2 NA POPOV
57STP 2 699 2 1 VE GOLANT
58STP 3 1223 8 1 VE MITSUK

309 2 H MARGENAU
56SET 3 14 8 1 GM PATEIUK
56SET 3 147 8 2 NA POPOV
57SET 3 895 3 1 AV GUREVICH
57STP 2 684 14 1 VE GOLANT

316 1 LM HARTMAN
56SET 3 14 8 1 GM PATEIUK

326 1 H MARGENAU
56SET 3 14 8 1 GM PATEIUK
57STP 2 1919 5 1 VE GOLANT

344 1 DE ALBURGER
60SET10 700 8 2 SA BEL'SKII

360 2 HB CASIMIR
56SET 2 73 1 1 EM LIFSHITZ
57SET 3 819 8 2 BV DERIAGIN
57SET 3 977 1 1 IE DZIALOSHINSKII

383 2 H LEVINE
58STP 3 740 12 2 NN LEBEDEV

405 4 NP HEYDENBURG
63SET16 47 2 1 AM PETROV

407 1 J SCHWINGER
57SET 4 259 1 1 AI BAZ'
57SET 4 612 1 2 IA ALEKSANDROV
57SET 5 144 3 3 VS BARASHENKOV
58SET 6 228 5 1 IA ALEKSANDROV

FIG. 1. Excerpt from Citations of Physics Literature, as Published.

58SET 7 794 7 1 LI LAPIDUS
60SET10 1281 5 1 PS OT-STAVNOV
60SET11 459 1 1 VM KOPROV
61SET13 1319 1 3 YA ALEKSANDROV
63SET16 1321 8 3 SB GERASIMOV

420 4 LS SKAGGS
56SET 3 663 3 1 KN SHLIAGIN

449 2 E FEENBERG
62SET14 393 3 1 AI NIKISHOV

470 1 PR WEISS
61SSS 2 2009 5 1 AM VASIL'EV

496 1 DQ POSIN
56SET 2 533 3 2 GS SOLNTSEV
57STP 2 684 2 1 VE GOLANT
58SET 7 693 2 1 GA ASKAR'IAN

518 1 DL FALKOFF
57SET 3 866 2 3 IV ESTULIN

521 1 KI GREISEN
60SET10 637 12 3 VA DMITRIEV

534 4 C TURNER
58STP 3 1420 5 2 LI PIVOVAR

549 1 F SEITZ
57SET 5 707 8 1 VS MASHKEVICH
57STP 2 237 9 1 TA KONTOROVA
58STP 3 443 4 1 VA CHUENKOV
58STP 3 876 11 2 MA KRIVOGLAZ
58STP 3 2194 3 1 GE PIKUS
59SET 9 1237 1 1 VS MASHKEVICH
60SET11 1144 4 2 ON KROKHIN
61SSS 2 1664 9 1 VA KOSTRYGIN
61SSS 2 2354 7 1 VA YAKOVLEV
63SSS 4 2336 4 1 VV RUMYANTSEV

601 2 CS COOK
57SET 4 355 12 1 NS SHIMANSKAIA

679 3 N BLOEMBERGEN
55SET 1 37 10 1 SA AL'TSHULER
56SET 2 486 7 2 SD GVOZDOVER
57STP 2 314 3 3 IS SHPIGEL'
58SET 7 695 3 1 AI RIVKIND
58STP 3 125 13 2 IN MOLIN
59SET 8 661 1 2 SA AL'TSHULER
59SET 8 720 11 1 GR KHUTSISHVILI
59SET 9 335 6 2 GV SKROTSKII
59SET 9 353 1 1 AA KOKIN
59SET 9 1242 2 1 KA VALIEV
60SET11 243 10 2 PG TISHKOV
60SET11 883 1 1 KA VALIEV
61SET13 903 11 3 AG LUNDIN
61SET13 1211 3 2 IV ALEKSANDROV
61SSS 2 1653 8 1 UK KOPVILLEM
62SET14 1116 1 2 RK TIMEROV
62SET14 1126 1 1 BN PROVOTOROV
62SET15 353 1 2 KA VALIEV
63SET16 822 3 1 NS KUCHERYAVENKO
63SET16 1269 2 2 IP IPATOVA
63SSS 4 1633 3 2 JS WAUGH
63SSS 4 2155 11 1 BN PROVOTOROV

718 2 JE NAFE
58SET 6 820 5 1 VM KONTOROVICH
59SET 9 120 7 2 MF DEIGEN

729 4 HA FAIRBANK
56SET 3 568 10 2 BN ESEL'SON

742 3 WW CHUPP
57SET 4 923 2 1 LP SOLOV'EVA

749 1 A RADKOWSKY
59SSS 1 888 8 1 I PASTRNYAK
61SSS 2 2637 26 2 EF GROSS

762 2 WG POLLARD
58SET 7 1016 2 1 O GHERMAN
62STP 7 353 9 1 VS TROITSKII
63STP 7 745 5 1 VI LOZGACHEV

CITATIONS OF PHYSICAL REVIEW

VOL. 73 (1948)

```
16   1 TP KOHMAN
  62SET14 1265    3 2 KA PETRZHAK 138 S6(1)

30   1 B  HOFFMANN
  57SET 5  642   19 1 A  POPOVICI 45 S5(5)

46   3 L  APKER
  56SET 3  159   17 2 ND MORGULIS 1238 J4(2)
  58STP 3 1752    4 2 PG BORZIAK 138 N8 S1(1)

51   1 GC WICK
  57SET 5  323    6 2 DF ZARETSKII 45 M4(5)

83   2 L  GOLDSTEIN
  61STP 6   38    4 4 VE GOLANT 1358 P1*

111  4 DJ HUGHES
  56SET 2  301    3 3 LE LAZAREVA 1238 S13(1)

127  3 HW LEWIS
  61SET12  527    5 2 VM MAKSIMENKO 45 Q4*
                        S4(5)
  61SET12  554    1 2 BT VAVILOV 45 Q4*

155  1 C  KITTEL
  56SET 2  703    6 1 AI PIL'SHCHIKOV 58 L5
                        N5 P5*
  58SET 6 1155   12 1 VL GUREVICH 45
  60SSS 2  332    4 1 SV TYABLIKOV 5 M4(5)
  61SET12 1111    1 2 AS SHUR 138 04
  61SSS 2 1805   12 1 SV TYABLIKOV 5 S5(5)
  62SSS 4 1165    8 1 EN YAKOVLEV 4568 M5
  62STP 7  316    1 2 DI MASH 1358 M4(35)
  63SSS 4 1282    5 1 VL BESHIDZE 345 M5
  63STP 7  820    6 2 DI MASH 12358 M4(2)

162  2 C  KITTEL
  55SET 1   37    8 1 SA AL'TSHULER 458 M5(5)
  61SSS 2 2009    2 1 AM VASIL'EV 5 M45(5)
  62SSS 4  175    3 3 VK KONYUKHOV 1238 M5(2)

241  4 NP HEYDENBURG
  62SET15  243    1 4 IB TEPLOV 138 N8 S1(1)
  63SET16   47    2 1 AM PETROV 138 L1(1) N8

256  4 CT LANE
  57SET 5  819    5 1 VM KUZNETSOV 158 08

257  2 EP HINCKS
  63SET16 1073    1 1 B  PONTECORVO 4568 J1

262  2 E  MCMILLAN
  55SET 1  576    1 2 PS BARANOV 18 M1(8)
  56SET 2  677   13 3 VI GOL'DANSKII 13468
                        S1(1)

266  1 PV HOUGH
  59SET 9 1263   17 2 AA VARFOLOMEEV 25 M1(5)

268  2 BT MATTHIAS
  60SC  4   54   11 2 EG FESENKO 18 L1(1) N8

297  1 H  MARGENAU
  55SET 1  205    3 1 VM FAIN 5 K5 S5(5)
  56SET 3   14    8 1 GM PATEIUK 18 S4(1)
  56SET 3  147    7 2 NA POPOV 18 S4(1)
  57STP 2  699    2 1 VE GOLANT 12358 M4(5)
  58STP 3 1223    8 1 VE MITSUK 1358 M5(8)

309  2 H  MARGENAU
  56SET 3   14    8 1 GM PATEIUK 18 S4(1)
  56SET 3  147    8 2 NA POPOV 18 S4(1)
  57SET 3  895    3 1 AV GUREVICH 45 S5(5)
                        T5(5)
  57STP 2  684   14 1 VE GOLANT 1456 M4(5)

316  1 LM HARTMAN
  56SET 3   14    8 1 GM PATEIUK 18 S4(1)

326  1 H  MARGENAU
  56SET 3   14    8 1 GM PATEIUK 18 S4(1)
  57STP 2 1919    5 1 VE GOLANT 35 M5(5)

344  1 DE ALBURGER
  60SET10  700    8 2 SA BEL'SKII 1358
                        M1(3)

360  2 HB CASIMIR
  56SET 2   73    1 1 EM LIFSHITZ 45 05
                        S5(5)
  57SET 3  819    8 2 BV DERIAGIN 458 J45
  57SET 3  977    1 1 IE DZIALOSHINSKII 45
                        N4 Q5*

383  2 H  LEVINE
  58STP 3  740   12 2 NN LEBEDEV 45 M4 Q4*

405  4 NP HEYDENBURG
  63SET16   47    2 1 AM PETROV 138 L1(1)
                        N8
```

FIG. 2. Excerpt from Citations of Physics Literature, with Citation Relationship Indicators Added.

407 1 J SCHWINGER
57SET 4 259 1 1 AI BAZ' 45 P5*
57SET 4 612 1 2 IA ALEKSANDROV 18 S4(1)
57SET 5 144 3 3 VS BARASHENKOV 45 M4
58SET 6 228 5 1 IA ALEKSANDROV 1238 L4(4) M45(5) S4(5)
58SET 7 794 7 1 LI LAPIDUS 345 M5
60SET10 1281 5 1 PS OT-STAVNOV 138 L4(1) M4(5)
60SET11 459 1 1 VM KOPROV 5 M4 Q5*
61SET13 1319 1 3 YA ALEKSANDROV 18 L6(1)
63SET16 1321 8 3 SB GERASIMOV 45 L4(4) Q4*

420 4 LS SKAGGS
56SET 3 663 3 1 KN SHLIAGIN 18 J1

449 2 E FEENBERG
62SET14 393 3 1 AI NIKISHOV 456 M1(5)

470 1 PR WEISS
61SSS 2 2009 5 1 AM VASIL'EV 5 M5

496 1 DQ POSIN
56SET 2 533 3 2 GS SOLNTSEV 138 K8
57STP 2 684 2 1 VE GOLANT 1456 J1
58SET 7 693 2 1 GA ASKAR'IAN 58 M5

518 1 DL FALKOFF
57SET 3 866 2 3 IV ESTULIN 128 05

521 1 KI GREISEN
60SET10 637 12 3 VA DMITRIEV 1238 M4(2)

534 4 C TURNER
58STP 3 1420 5 2 LI PIVOVAR 138 N8

549 1 F SEITZ
57SET 5 707 8 1 VS MASHKEVICH 45 M4(5)
57STP 2 237 9 1 TA KONTOROVA 35 M4(5)
58STP 3 443 4 1 VA CHUENKOV 345 M14(5)
58STP 3 876 11 2 MA KRIVOGLAZ 45 L4(5)
58STP 3 2194 3 1 GE PIKUS 5 L47
59SET 9 1237 1 1 VS MASHKEVICH 345 M4
60SET11 1144 4 2 ON KROKHIN 45 M4(5)
61SSS 2 1664 9 1 VA KOSTRYGIN 138 M4(3)
61SSS 2 2354 7 1 VA YAKOVLEV 235 M5(2)
63SSS 4 2336 4 1 VV RUMYANTSEV 3458 M4(3)

601 2 CS COOK
57SET 4 355 12 1 NS SHIMANSKAIA 15 L1(15) 08

679 3 N BLOEMBERGEN
55SET 1 37 10 1 SA AL'TSHULER 458 M5
56SET 2 486 7 2 SD GVOZDOVER 1458 08 S1(1)
57STP 2 314 3 3 IS SHPIGEL' 5 M47(5)
58SET 7 695 3 1 AI RIVKIND 234 08
58STP 3 125 13 2 IN MOLIN 1238 M5(2)
59SET 8 661 1 2 SA AL'TSHULER 3458 04
59SET 8 720 11 1 GR KHUTSISHVILI 458 M4(5)
59SET 9 335 6 2 GV SKROTSKII 345 L4 N4 P5*
59SET 9 353 1 1 AA KOKIN 45 K4
59SET 9 1242 2 1 KA VALIEV 4568 J4 Q5*
60SET11 243 10 2 PG TISHKOV 123 M4(3)
60SET11 883 1 1 KA VALIEV 345 Q5*
61SET13 903 11 3 AG LUNDIN 138 M4(3)
61SET13 1211 3 2 IV ALEKSANDROV 45 M5
61SSS 2 1653 8 1 UK KOPVILLEM 345 M5
62SET14 1116 1 2 RK TIMEROV 45 M5 Q4* S4(5)
62SET14 1126 1 1 BN PROVOTOROV 345 N4 Q4*
62SET15 353 1 2 KA VALIEV 45 Q4*
63SET16 822 3 1 NS KUCHERYAVENKO 1238 M7
63SET16 1269 2 2 IP IPATOVA 45 J4
63SSS 4 1633 3 2 JS WAUGH 345 N45 P5*
63SSS 4 2155 11 1 BN PROVOTOROV 45 05 T5(4)

718 2 JE NAFE
58SET 6 820 5 1 VM KONTOROVICH 45 M7
59SET 9 120 7 2 MF DEIGEN 456 M1(6)

729 4 HA FAIRBANK
56SET 3 568 10 2 BN ESEL'SON 18 P8*

742 3 WW CHUPP
57SET 4 923 2 1 LP SOLOV'EVA 138 J1

749 1 A RADKOWSKY
59SSS 1 888 8 1 I PASTRNYAK 1238 M4(3)
61SSS 2 2637 26 2 EF GROSS 12358 S4(1)

762 2 WG POLLARD
58SET 7 1016 2 1 O GHERMAN 345 N4
62STP 7 353 9 1 VS TROITSKII 345 L5
63STP 7 745 5 1 VI LOZGACHEV 45 S5(5)

which appear to characterize much of the activity of physicists: data transformation (as distinct from primary observation); and calculation from theory (as distinct from formulation of basic concepts of new theory). The value of this group of categories as applied to a large citation index is that it tells the user who is interested in working outward from the cited references specifically in one of the (at least) eight possible directions whether there is any chance whatever of finding pertinent information in each citing reference. For example, if he is only looking for new experimental data, he can quickly rule out a citing reference which it indicated to be purely theoretical.

Group Two includes categories which describe *other-than-original scientific contributions* of the citing publication. It can be thought of as describing the administrative or associative relation to science in general—the secondary effort to organize earlier scientific contributions into useful packages. Three categories are included in the group at present: review article; bibliography; data cumulation. For a scientist in search of summarized or cumulated information, these three categories may prove sufficient. It is easy to see, however, that these categories of secondary contribution can well be fragmented into many subcategories which could be valuable to librarians and research administrators.

Group Three consists of categories which describe the *continuity relationship or identity relationship* between the citing paper and a specific cited paper. (Group One and Group Two categories would be identical for each listing of a citing reference in a citation index, while Group Three and Group Four references would vary with the item cited.) There are seven categories in Group Three at present, each indicating a different kind of author-stated, or publisher-stated, continuity relationship between the cited paper and the citing paper which is independent of the scientific or secondary contribution. The categories are: shared authorship (one or more authors in common on citing and cited publications); same text (different versions published separately); abstract or condensation; erratum; continuation; precursor (citing paper refers to a publication which will follow); inclusion (citing publication is a composite which contains the cited paper, e.g., a separately authorized chapter in a book). The categories in this group are of potential value to citation index users in two ways. The identity indicators (same text; abstract or condensation; inclusion) tell the user where he may find a potentially satisfactory substitute for a known reference which may be unavailable to him in the original. The continuity indicators (shared authors; erratum; continuation; precursor) designate references which definitely or very likely contain information which will amplify or qualify the specific work reported in the known reference.

Group Four consists, at present, of eleven categories which indicate a *disposition relationship*—what the citing publication has done about the cited publication. The categories of disposition are: noted only (the existence of the cited paper is mentioned but nothing further is said or done about it); distinguished (the cited paper is mentioned as being not pertinent to the subject of the citing paper); reviewed or compared; applied; improved or modified (as a theory or an experimental technique); replaced (with an alternative theory or method); changed the precision of the cited work (plus or minus); changed the scope of applicability (plus or minus); questioned (expressed doubt); confirmed or affirmed; refuted. In assigning Group Four indicators, it is convenient to do so in combination with indicators from Group One in order to specify precisely which scientific aspect of the cited paper was disposed of by the citing paper and in what manner. For example, one can indicate by a short three-symbol code that observed data from the cited paper was affirmed by the citing paper by means of logical explanation. This provides considerable information on the nature of the citing reference even though no subject terms are mentioned; the subject matter may be inferred with fair accuracy by the index user because he is familiar with the cited publication.

Analysis of actual publications for the purpose of determining relationships according to the scheme described above is neither simple nor rapid. As would be expected, however, it tends to become faster and easier with practice. Reproducibility of analyses (when a paper is analyzed a second time some weeks after the first analysis) has been well under 100%; nevertheless, the results have been very similar and there are indications that individual differences in the assignment of relationships can be largely overcome by the application of somewhat broadened strategies in index use. For example, it is entirely likely that a paper which is indexed only for a scientific contribution of description (experimental observations) will contain a few details on experimental technique even though the indexer may not have considered them sufficiently enlightening to warrant an indicator for experimental technique. Similarly, one should recognize that the borderline between prediction and calculation from theory is fairly arbitrary to persons who are facile in mathematics. Thus, by widening one's interest to include the categories which are most likely to be confused with the categories of principal interest, one can deal with the differences among indexers. This is completely analogous to the way in which users currently cope with differences among subject indexers.

It is planned to broaden this work in the future to gage the relative usefulness of the different relationship indicators to individual research physicists. It is also planned to investigate the problems of training indexers to recognize the various categories of relationship. Work in other fields of science is contemplated also. It is of particular interest to determine whether scientists themselves can be taught to apply this kind of analysis to publications.

Citing journal

- SC: Soviet Physics—Crystallography
- SET: Soviet Physics—JETP
- SSS: Soviet Physics—Solid State
- STP: Soviet Physics—Technical Physics

Contribution

- 1: Observed phenomena, description, experimental data.
- 2: Data transformation.
- 3: Explanation (logical deduction of observed data from theory and premises).
- 4: Theory, hypothesis.
- 5: Calculation from theory.
- 6: Prediction of observable phenomenon.
- 7: Notation, definition.
- Experimental method, manipulative prescription.

Reason for citing (operational relationship of citing paper to cited paper)

- J: Noted only—no clear indication of reason.
- K: Distinguished from the primary focus of citing paper.
- L: Compared, reviewed, re-examined.
- M: Applied.
- N: Modified.
- O: Replaced with alternative.
- P: Achieved different precision: greater (*) or less (—).
- Q: Achieved different scope of applicability: greater (*) or less (—).
- R: Questioned.
- S: Affirmed.
- T: Refuted.

FIG. 3. Key to Codes.

If they can, then it is conceivable that future authors can be induced to include relationship indicators along with the bibliographic references in their publications; this would have the happy effect of making possible the production of citation indexes of high selectivity without requiring the employment of indexers, since the relationship indicators would merely have to be copied along with the citations.

• Illustration of a Citation Index Incorporating Relationship Indicators

An illustration is included to show how a citation index to science literature might be improved through the addition of indicators of citation relationships. In Fig. 1 is presented a short excerpt from an experimental citation index to a small body of physics literature.[5] This index is arranged for entry by journal, volume, and page of the cited reference. Figure 2 shows the same excerpt, but with citation relationship indicators added after each citing reference. The key to codes is given in Fig. 3, and a sample index entry is interpreted in Fig. 4. While the excerpt contains few cases where multiple citing might

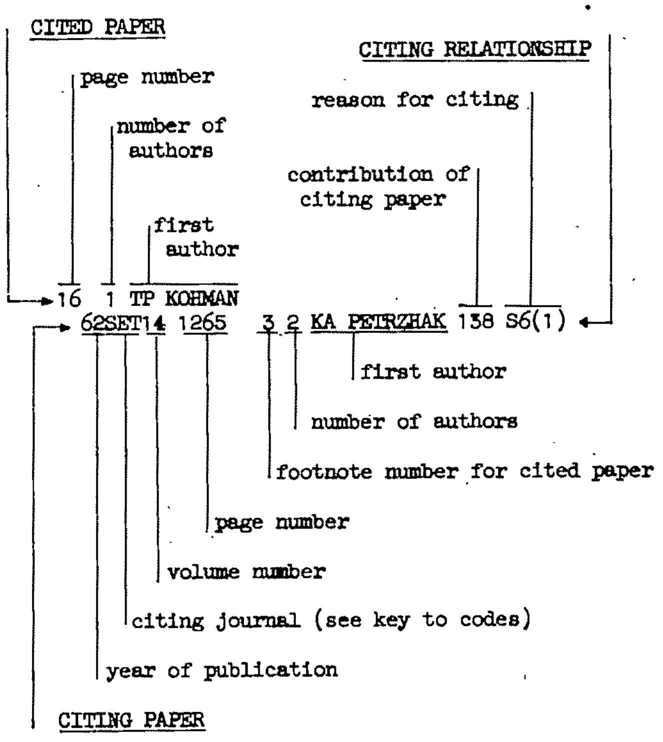


FIG. 4. Illustrative Index Entry.

INTERPRETATION: The paper by T. P. Kohman on page 16 (of Volume 73 (1948) of *The Physical Review*) was cited in footnote 3 of a paper on page 1265 of Volume 14 (1962) of *Soviet Physics—JETP* by K. A. Petrzhak and one co-author. The Petrzhak paper contributes experimental data (1), explanation of experimental data (3), and description of experimental method (8); it confirms (S) a prediction (6) in the Kohman paper with experimental data (1).

constitute a serious look-up problem for the user (because the index coverage is small), the relationship indicators in Fig. 2 nevertheless afford the user a definite opportunity for time saving through elimination of references which cannot satisfy his requirements. With a citation index drawn from a large body of literature, the opportunities for time saving would be much more dramatic.

This illustration of the use of relationship indicators contains only GROUP ONE and GROUP FOUR indicators. Its purpose is purely to elucidate an approach. It is not suggested that the GROUP TWO and GROUP THREE indicators should be dispensed with; they merely happened to be inapplicable to the body of literature in the illustration. The coding scheme for the relationship indicators is illustrative only and not necessarily intended as a standard to be followed in future work along this line.

References

1. B. A. Lipetz, "Design of an Experiment for Evaluation of the Citation Index as a Reference Aid," *Automation and Scientific Communication,* Preprints of the Annual Meeting of the American Documentation Institute, Chicago, October 6–11, 1963, pp. 265–266.
2. A second important drawback is that information may be lost to the index user if citations are merely copied from the source literature without any effort at verification of the accuracy of the cited reference. Published references often contain typographic errors and bibliographic inconsistencies[3] which affect the automatic sorting of index entries and which result in entries being published incorrectly in index positions where the user would not look for them.
3. B. A. Lipetz, "Compilation of an Experimental Citation Index from Scientific Literature," *American Documentation,* Vol. 13, 1962, pp. 251–266.
4. B. A. Lipetz, "An Approach to the Measurement of Efficiency of Scientific Research Organizations" (Doctoral thesis), Cornell University, June 1959, 317 pp.
5. B. A. Lipetz, assisted by C. F. Dziurkiewicz, *Citations of Physics Literature, Experimental Issue,* New York: American Institute of Physics, 1964, 105 pp. (AIP/DRP–C1 1964). Excerpt is from pp. 23–24.

Random Code Numbers for Universal Identification of Documents

C. W. McCUTCHEN

Section on Rheumatic Diseases
Laboratory of Experimental Pathology
National Institutes of Health
Bethesda, Maryland

• The Problem

Much effort is being expended on making human knowledge accessible to human beings. The emphasis is on methods of classifying and indexing documents so that a search by subject can be speedy and thorough.

I suggest that this is only part of the problem, the other part being to find documents whose titles one already knows. All of us must remember frustrating searches in library catalogs which ended neither with the retrieval of the document nor with the certainty that it was not in the library.*

This happens most often when we search for an *obscure* document, the elements of obscurity being, in descending order of importance: 1. the absence of any visible author; 2. a long title, laden with information; and 3. remoteness from the library's field of specialization. Typical examples would be conference reports, preprints, privately circulated reports from commercial and government laboratories; in fact, almost anything which is neither a text nor a journal.

Undoubtedly, many libraries deal effectively with the examples mentioned, but their success is evidence of their skill (and perhaps affluence), not of the unimportance of the problem. In contrast, the difficulties posed by obscure documents are confirmed by the reluctance of many libraries to give them house room, a phenomenon which is doubly unfortunate because we lose the information that the documents contain, yet are hardly aware that the loss is occurring. We simply do not expect to find these things in libraries.

* In this article, a document is defined as the physical entity one searches for and removes from the library shelf. Thus, a conference report is a document, while a single article within it is only part of a document.

• Obscurity Explained

What makes obscure documents obscure?

When I look in the telephone book under John Y. Doe, his number or its absence is quickly found. People are much easier to look up than documents. There are no obscure people (in my limited sense).

The reason for the difference is that people have names and documents do not. Documents have titles, but titles and names are not necessarily the same thing.

My name serves one purpose only, that of identifying me. My name is Charles W. McCutchen, a fact which I can reveal with impunity because it tells nothing about me except where my ancestors came from. All it says, and this it says very clearly, is that I am not William Q. Smith or John J. Jones.

Further, there is a convention that names shall be listed alphabetically in the order of surname, first and middle. And it is precisely because there is no descriptive information in names that we succeed in sticking to this convention. No one will list me under "Charles" in order to group me with others possessing the trait of *charlesness*.

Document titles are chosen so as to describe the documents, exactly what names do not do. Of course, as they are not all identical, they can function as names, but they do it very badly.

To begin with, they are inefficient. Far too many start with stylized sequences like "An introduction to . . ." Were that all, one could still list them alphabetically and the library user would find his document every time, though the decisions which separated it from its neighbors in the catalog would often occur at the twentieth letter or beyond.

Far worse than inefficiency is information content, because the user tends to remember the information and forget the exact wording.

Most libraries try to get over these difficulties by leaving out the minor words and/or rearranging the title is a systematic manner. These are procedures which simplify searching for standardized documents like texts and journals, but which lack the utter predictability of pure alphabetic listing. Thus, the cataloging of obscure documents comes to depend upon a complicated set of rules which the library user seldom knows, and which it is unreasonable to expect him to learn or remember.

Nothing else so clearly demonstrates how poorly titles perform the identifying function as the almost universal preference for the author catalog. We use the author's name as a stand-in for the name which the document itself does not possess. Alas, not all documents have visible authors.

The contrast is startling when we visit the fiction section of the library. To know the title is to find the book. And for good reason. Titles of novels are more nearly names than descriptions of content. The words *Wuthering Heights* do little more than suggest a mood. But suppose Miss Brontë had regarded her work as a sociological treatise. It might have been called *A Study of Upper Class Life in England During the Industrial Revolution.* Imagine cataloging that so I could find it!

Briefly, one can have descriptive titles, or one can have names, but one cannot get both services effectively from the same set of words.

• The Solution

With novels, one set of words is enough. The title does not try to be a description, but other documents need both a descriptive title and a name.

And why not? If all documents were christened at publication they could be cataloged by their names and even people like me could find them. The names need have no sales appeal. They need not be euphonious. They need not even be pronounceable.

They should be efficient and, most important, they should be meaningless, because only this enforces a unique (in the sense of one and one only) system of cataloging. Finally, and perhaps almost as important, their selection must require little effort from the publisher.

These requirements will all be met if, *to the title of each document there be appended, on the document itself, a meaningless code number or code word made up of digits or letters chosen at random by the publisher.*

Making up a code number or word takes about two man-minutes and a pair of dice (see Appendix 1 for details), and less traditional methods would no doubt be quicker, if not so interesting.

The relative merits of digits and letters are discussed in Appendix 2. Both have advantages. I incline slightly toward digits and will henceforth assume their use, but any conclusions will, with obvious translation, hold equally well for letters.

To visualize the system in operation and to try to understand the problems which will occur during the transition period, we shall do violence to history and imagine that code numbers were introduced in 1900:

At first their spread was slow: Publishers saw no reason to put on the numbers unless libraries had code number catalogs, and librarians were unwilling to start new catalogs just because a few documents carried numbers and a few enthusiasts included them in reference citations

It was the airplane which gave them their real start. With the airplane came government aerodynamics laboratories, and from these laboratories flowed, from 1910 on, a steadily increasing stream of technical reports, notes, and memoranda. Probably some aerodynamicist was one of the believers, but they all knew each other, and just who was the moving spirit is now forgotten. It is, however, an observable fact that after 1920 every one of these documents was coded.

Once really started, their use spread rapidly, with the publishers (perhaps prodded by the authors— also see below) taking the lead. By 1930, code numbers appeared consistently on everything except novels.

Aerodynamics libraries at first, and then reference libraries generally, code-cataloged what they could, but until all citing authors learned to include the code number, every document still had also to be entered by title. By 1935, citation by code number had become universal, and in many libraries the title catalog either dropped out of use (except for older documents) or merged with the subject catalog.

Even now, not all authors cite by code number alone. Many include the descriptive title as well because they feel that it helps their readers to decide which references are of interest.

Recently a new scheme has been proposed. The idea is to put an enormous number of documents on microfilm or video tape at a few central libraries. Readers at subscribing libraries would be able to dial the code number of any document, and the page number, and then either read the page by television or have it printed out by facsimile. This should put each reader in potential touch with 30 million documents. Few of the subscribing libraries could afford a catalog covering even a small fraction of these, but all are available to the user who knows the code number. As these are now included in all citations, research workers should find the absence of a catalog a bearable inconvenience.

This and other proposed forms of library automation have got the fiction publishers looking to the future, and quite a few novels now appeal with code numbers.

In historical perspective, it is easy to see the virtues of code numbers, but it was not always so obvious. Some people worried that because code numbers were picked at random some of them might be used twice. Why they had not been equally afraid that authors of the same name

might write works with the same title is not clear, but just to be on the safe side (Appendix 3) the originators of the idea suggested that the numbers be fifteen digits long, split into groups of three, like this: 012-015-430-229, so as not to be confusing. They are still printed that way, though users of small libraries ignore all but the first six digits. Occasionally they get two titles to choose from, but they do not seem to mind.

Another early objection was that random numbers are hard to remember. This their protagonists admitted, but they countered with the argument that people who used reference citations seldom tried to remember them, so the effect of random numbers was merely the substitution of one form of jotted-down note for another. The point appears to have been well taken, because the controversy died down as the spread of the system allowed the objectors to experience its operation.

At the beginning, some publishers used code words instead of numbers, and libraries had catalogs for each. This caused no real trouble, but the code words were unpopular with the readers and their proponents soon switched to numbers. Big libraries still have one or two catalog drawers of code word entries, but they are all old and seldom referred to.

Also at the beginning, some of the journal publishers assigned a new number to each volume, and some even to each issue. It worked well, but libraries complained about the number of cards, and the practice soon stopped.

At times there have been proposals for other kinds of code numbers. Usually the idea was that information of one kind or another would be encoded as part of the number, and the rest would be a serial number assigned by a central authority. It was advanced that the information would be useful, and that the shortness of the serial number would be an advantage.

The random number supporters answered back that codes which contained information could be cataloged according to that information, and sometimes would be, and where would that leave the library user? And while admitting that random numbers are necessarily longer than serial numbers because they achieve uniqueness by statistics rather than assignment, they pointed out that the part that actually has to be used on most occasions is the same length (Appendix 3).

Perhaps these arguments helped, but what really defeated the other proposals was that no two groups wanted to have the same information encoded, nor was it ever settled who would assign the serial numbers.

These proposals were frequent all during the twenties, and it is interesting to note that this was the period when the bulk of publishers overcame their reductance to put on the random numbers.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	~~4~~	~~5~~	~~6~~	~~7~~	8
3	4	~~5~~	~~6~~	~~7~~	~~8~~	9
4	5	~~6~~	~~7~~	~~8~~	~~9~~	10
5	6	~~7~~	~~8~~	~~9~~	~~10~~	11
6	7	8	9	10	11	12

FIG. 1. Latin Square for Generation of Random Numbers by Rolling Dice.

• Appendix 1

Generation of Random Characters by Rolling Dice

For digits, roll an ordinary pair of dice. Ignore those throws which turn up neither a one nor a six. On the remaining throws, add up the spots and take only the last digit (10, 11, and 12 are recorded as 0, 1, and 2 respectively). The Latin square of possible combinations (Fig. 1) shows that there are two ways of rolling each digit and each is thus equally likely.

For letters, the easiest method requires two dice which one can tell apart. One rolls the dice and uses the Latin square shown in Fig. 2 to assign letters to the possible combinations.

Rolling dice may not be in the spirit of modern technology, but the characters generated are as random as the dice and their shaking are honest. Computer techniques, unless they involve a noise generator or other randomizing device, produce sequences of pseudo random numbers which are related to each other in an involved but perfectly definite manner, and considerable care is required to ensure that such techniques are equally likely to generate all the possible permutations. If they do not, then the possibility of multiple use of the same code numbers is increased. In particular, there is the danger that two or more publishers will use the same computer routine and the same input numbers, and produce long identical sequences.

The likelihood of this is not easy to estimate—while dice, we know, are safe.

	1	2	3	4	5	6
1	a	b	c	d	e	
2	f	g	h	i	j	
3	k	l	m	n	o	
4	p	q	r	s	t	
5	u	v	w	x	y	
6						z

FIG. 2. Latin Square for Generation of Random Letters by Rolling Dice.

• Appendix 2

Relative Advantages of Digits and Letters

A code word is only 5/7 as long as a code number which provides the same number of permutations, and separation of a document from its neighbors in a catalog requires that one look at only 5/7 as many characters.

Yet, use of letters may not lead to easier searching, because it is doubtful if many people remember the order of the alphabet in the automatic manner in which they know the order of the digits. The order of the alphabet is an arbitrary convention. The digits are ordered according to the size of the number each represents; which is again a convention, but one which is inevitably learned along with arithmetic.

Also, thanks to arithmetic, people are used to writing and reading all possible combinations of numbers.

The alphabet transcribes somewhat more efficiently into binary representation than does scale of ten, but equipment may be more readily available for the latter transcription.

Digits might tempt publishers into using tables of random numbers (a major blunder if the tables are used naively), but they remove the temptation to use the firm's initials, and thus make its products slightly harder to tell apart from each other.

And lastly, digits are completely international, nor, to my knowledge, does any culture regard any combination of them as more than mildly obscene.

• Appendix 3

Multiple Assignment of Random Code Numbers

Assigning random code numbers to documents is the same, in its effect, as scattering documents at random into a vast array of boxes where each box represents a possible code number.

Confusion arises when two or more documents land in the same box and receive the same code number. So, given u boxes into which we distribute n documents, how many boxes will be doubly occupied, how many triply occupied, and so on?

To find out, we fix our attention on one box and calculate the likelihood of the various degrees of occupancy. Multiplying the resulting figures by u then gives the number of boxes with each degree of occupancy averaged over many repetitions of the scattering experiment.

The *a priori* probability that a document will fall into a particular box is $I = \frac{1}{u}$ and the probability that it will fall elsewhere is, of course, $E = 1 - \frac{1}{u}$. When n documents are distributed the probabilities that there will be 0, 1, 2, 3, etc., documents in our particular box are given by successive terms in the binomial expansion

$$(E - I)^n = E^n + nE^{n-1}I + \frac{n(n-1)}{2}E^{n-2}I^2 \ldots I^n$$

or, in general, if we let k be the number of documents in a box, the probability P_k of this degree of occupancy is given by

$$P_k = \frac{n!}{k!(n-k)!}E^{n-k}I^k$$

and the average number N_k of boxes with this degree of occupancy

$$N_k = uP_k = \frac{un!}{k!(n-k)!}E^{n-k}I^k = \frac{n!}{k!(n-k)!}u^{1-k}\left(1-\frac{1}{u}\right)^{n-k}$$

This is more usable if we take $\log_{10}$ of both sides, getting

$$\log N_k = \log\frac{n!}{k!(n-k)!} + (1-k)\log u + (n-k)\log\left(1-\frac{1}{u}\right) \quad (1)$$

Because we are only interested in cases where $u \gg 1$ we can approximate $\log\left(1-\frac{1}{u}\right)$ by $-\frac{1}{u}\log_{10} e = -\frac{.434}{u}$. Also, because we are only interested in values of $k \ll n$, we replace $\frac{n!}{(n-k)!}$ by n_k *and* $n\text{-}k$ by n, getting

$$\log N_k = k\log n - \log k! + (1-k)\log u - .434\frac{n}{u} \quad (2)$$

The number of doubly and more highly assigned code numbers will be small only if $n \ll u$, so we introduce small error by dropping the last term, which gives

$$\log N_k = k\log n - \log k! + (1-k)\log u \quad (3)$$

from which we deduce immediately that N_k is proportional to n^k and inversely proportional to u^{k-1}.

If we let $u = 10^{15}$ and $n = 10^7$ then we find that the number of

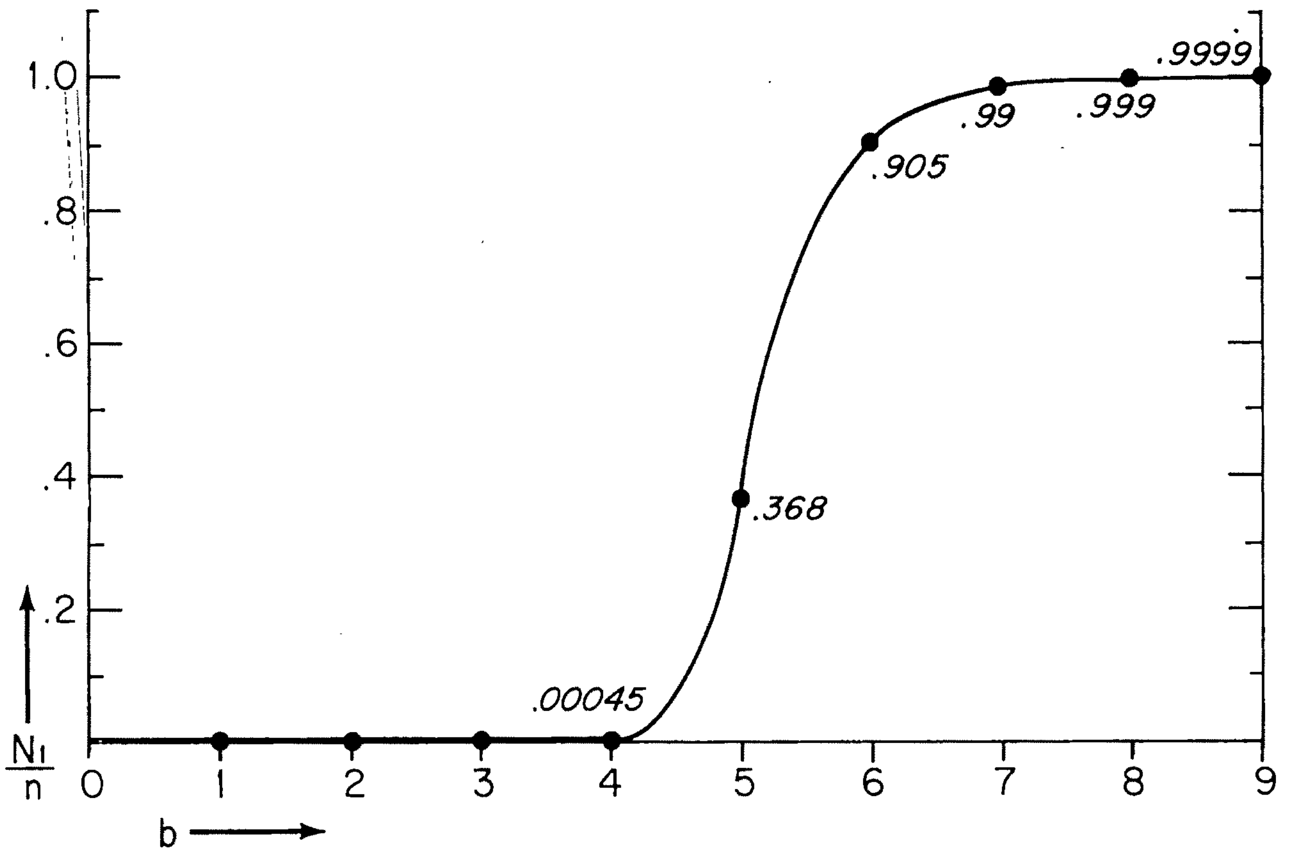


FIG. 3. The Fractional Integral Curve of Separation (Eq. 5) Plotted for $n = 10^5$.

doubly assigned code numbers $= N_2 = .05$
triply assigned code numbers $= N_3 = 1.67 \times 10^{-10}$
quadruply assigned code numbers $= N_4 = .416 \times 10^{-18}$

so, were the entire content of the Library of Congress code cataloged with fifteen digit numbers, it is unlikely that there would be a single doubly assigned number. Higher assignments clearly need cause no concern.

Were the content of the library to increase to an astronomical 10^{11} documents, there would be only 5×10^6 doubly assigned numbers. The use of *only* to describe five million may occasion surprise, but it means that only once in twenty thousand retrievals would an extra document appear along with the one desired, a rate of false retrieval far less than already occurs for other reasons.

Cataloging Efficiency of Random Code Numbers

To how many digits of a random code number will one have to refer, on the average, to distinguish it from its neighbors in the catalog?

If we look only at the first digit the chance of distinguishing a document from its neighbors is minute. We are, in effect, attempting to catalog with one-digit code numbers, and the chance that our document will be singly assigned to a one-digit number is small. When we look at the second digit as well, we raise the code numbers we are actually using to two digits; there are more possible permutations so there will be more singly assigned numbers (though for a library of any size there will still be less than 1).

Proceeding further, there will eventually come a point where the number of single assignments increases sharply between successive digits, and it is in this range that we separate most of the documents.

The number of single assignments plotted against the number of digits examined can be called the integral curve of separation. It is an integral curve because its value at a particular digit gives the number of documents separated at that and at all previous digits.

To calculate this curve, we must go back to Eq. 2, because n is no longer always much smaller than u. For $k = 1$ this becomes

$$\log N_1^{-} = \log n - .434\frac{n}{u} \qquad (4)$$

It is easier to deal not in N_1, but in $\frac{N_1}{n}$, the fraction of the number of documents which have singly assigned code numbers. Thus,

$$\log \frac{N_1}{n} = -.434\,\frac{n}{u}$$

or

$$\frac{N_1}{n} = 10^{-.434\frac{n}{u}} = e^{-\frac{n}{u}}$$

If we are looking at the b^{th} digit then $u = 10^b$ and we can write

$$\frac{N_1}{n} = \exp\left(-\frac{n}{10^b}\right) \tag{5}$$

The fractional integral curve of separation, $\frac{N_1}{n}$ as a function of n, is plotted in Fig. 3 for $n = 10^5$, a size rather larger than the average reference library. Separation, of course, is done only at integral digits, so for $n = 10^5$ the curve has meaning only at these points. But it is all significant because should the library buy more books the curve would move bodily to the right, and the integral digit points would fall upon it at a new set of places.

Were the contents of our 10^5-volume library serially rather than randomly numbered, all separation would occur at the 4th digit, so the use of random numbers requires that one look, on the average, at an extra 1.6 digits.

But this situation is unrealistic. No library will contain all the coded documents, and few will contain even a fair fraction. As a library comes to hold a smaller and smaller fraction of all the coded documents the difference in operation between random and serial numbers becomes less and less. For example, should there be a total of 10^7 serial numbered documents, the separation curve would be the same as that for random numbers except that the part beyond $b = 6$ would be cut off, and the documents involved would be separated at digit 6 or lower.

Acknowledgment

It is a pleasure to acknowledge helpful criticism by Robert A. Fairthorne.

Improvements in a Permuted Title Index

Improvements in a computer-prepared, permuted-title index are described. The improvements include adding words to present titles or synthesizing new titles when insufficient keywords are provided in the title, positioning singular and plural words and word variants so that one follows the other, identifying and adding cross-references where needed, and exceptional efforts in removing words that have no value for information retrieval. The index was prepared for a radiobiology bibliography.

HUGH E. VORESS

Chief, Technical Literature Section
Reference Branch, Division of Technical Information Extension
Atomic Energy Commission
Oak Ridge, Tennessee

• Introduction

In 1960, the Division of Technical Information Extension (DTIE) of the U.S. Atomic Energy Commission was asked to prepare a bibliography of radiobiology information covering the years 1958, 1959, and 1960. The resulting efforts recently culminated in the issuance of an 1,824-page bibliography in three books, entitled *The Effects of Radiation and Radioisotopes on the Life Processes*. The first two books contain 11,944 annotated references.* Book three contains computer-prepared author and subject indexes. The bibliography bears the report number TID-3098.

Preliminary scanning of radiobiology literature indicated that this compilation would involve a large number of references. Inasmuch as sufficient man-hours were not available to prepare subject and author indexes by traditional manual methods, the decision was made to utilize computer techniques. After review of several computer-prepared indexes, it was further decided that, for these purposes, an improved permuted-title index would be developed that would resemble conventional book-type indexing. This was accomplished by adding cross-references, arranging correlative words together, and eliminating lines beginning with non-significant words.

This paper describes the steps employed for the preparation of indexes for this bibliography.

Before these steps are described in detail, however, it may be well to explain briefly the time required for each of the operations in preparing the complete bibliography.

* Major secondary reference sources used in locating this literature included: *The Bibliography of Agriculture, Biological Abstracts, Chemical Abstracts, Excerpta Medica, Index Medicus, International Abstracts of Biological Sciences, Journal of the Science of Food and Agriculture,* and *Nuclear Science Abstracts.*

The first year, beginning in April 1961, was spent in a literature search of seventeen abstract journals, preparing a subject classification scheme, eliminating duplicate references, and assigning subject categories to the references.

During the second year (April 1962-March 1963), the references were arranged in final order, given an abstract number, composed, proof read and corrected, and finally made up into book pages for offset printing. By the end of March 1963, all titles had been reviewed for permutation and approximately 20% of them had been altered to permit more specific indexing.

During October 1963, the keypunching of the titles, authors, and journals, and the first printouts of the subject and author indexes were completed. The final author index was received in November 1963. Modifications of the subject index described below were supplied for keypunching and merging into the index in January 1964.

The final printout of the subject index was received in February 1964, and offset printing was accomplished in April 1964.

In summary, approximately two years were devoted to preparing the text of the bibliography and one year to the indexes. It is estimated that it would have taken three man-years to prepare the subject index alone by traditional methods.

• Non-Significant Word Lists and Preparation of Titles

In the preparation of a permuted-title index, it has been found useful to identify words that are useless for information-retrieval purposes, and therefore should not appear as keywords. Our organization initially prepared

ABILITY	ACCORD	ACTIVE
ABOARD	ACCOUNT	ACT
ABOUT	ACCUMULATED	ACTS
ABOVE	ACHIEVED	ACTUAL
ABUSE	ACQUAINTED	ADDED
ACCESSIBLE	ACROSS	ADDENDUM
ACCOMPANIED	ACTED	ADDITIONAL
ACCOMPANYING	ACTING	ADDITION
ACCORDANCE	ACTION	ADDITIONS
ACCORDING	ACTIONS	

FIG. 1. Examples of Basic Words in DTIE Non-Significant Word List.

a list of 1,800 basic non-significant words, a portion of which is provided as Fig. 1. A review of the first printout of the permuted index produced additional basic words plus words considered non-significant for this index in radiobiology, although possibly meaningful in a permuted index in another subject field. A re-run of the permuted index eliminated these additional non-significant words from the final printed version.

The title for each referenced abstract was examined to determine whether adequate index words appeared. When a title was considered acceptable, the portion of the title to be keypunched for computer permutation was underlined. For example: "The Results of a Study of the Inherent Dangers of Radioactive Fallout in Agriculture." Inadequate titles were altered to incorporate additional index points as keywords suggested by the abstract, or else they were replaced by new, synthesized titles. The following example illustrates an inadequate title and some of the new titles as synthesized by the bibliography editors for permuting. The original title "The Seventh All-Union Congress of Roentgenology" was replaced and keypunched in each of the following ways to insure that adequate subject index points were created:

"Injurious Effects of Ionizing Radiation and Protective Measures Against It,"

"Effectiveness of Super-voltage X-radiation and Electron Beams in Therapy,"

"Pathological Effects of Chronic Exposure to Low Levels of Radiation,"

"Radioisotope Effectiveness in Cancer Therapy," and

"Hazards from Increases in Natural Background Radiation from Nuclear Weapons Tests."

In addition to the obvious problem cited in the above example, compound titles (consisting of main titles and subtitles) and progress reports for which titles were necessarily synthesized resulted in references being indexed by more than one title. Foreign language titles were necessarily translated. A rubber stamp code was used to indicate to the keypuncher when the original title was to be replaced by a synthesized title.

```
T1    6957         REACTION OF LIMB REGENERATES OF ADULT AXOLOTL
T2                 ( SIREDON MEXICANUM ) TO X-IRRADIATION
A3                 BRUNST VV
W4.   RADIATION RESEARCH 12, 642-56( 1960 ) JUNE.
T1    6958         LOCALIZED CHARACTER OF ULTRAVIOLET EFFECTS ON THE
T2                 URODELE FORELIMB
A3                 BUTLER EG + BLUM HF + SCHMIDT SE
W4.   J. CELL. AND COMP. PHYSIOL., 50 (3 ) 381-388( 1957 ).
T1    6959         EFFECT OF PARTIAL IRRADIATION ON THE MORPHOGENESIS OF
T2                 LIMB REGENERATES
A3                 GOSS RJ
W4.   J. MORPH., 101( 1 ). 131-148( 1957 ).
T1    6960         CAUSES OF REGENERATION ABILITY LOSS IN THE AXOLOTL
T2                 AFTER X-RAYING
A3                 SIDOROVA VF
W4    BULL. EXPTL. BIOL. MED. ( USSR ) ( ENGLISH TRANSLATION ) 43,
W5.    215-19( 1957 ) FEB.
```

FIG. 2. Printout of Unit Cards Showing Titles (T), Authors (A), and Journal Citations (W).

Three components of the descriptive cataloging (i.e., title, authors, and journal citation) were underlined in distinctive colors to guide the keypunchers. These components were coded separately as unit cards so that it would be possible to print out author and subject indexes and a list of the journals cited. An example of the actual keypunched information is shown in Fig. 2.

• Preliminary Permuted Index

An index was printed out in the form customarily published as a standard permuted-title index. However, this printout was regarded as preliminary inasmuch as it was believed that several improvements were both desirable and feasible. Therefore, this version was used as copy for the extensive editing necessary to produce a sophisticated index—one that would allow more complete and efficient information retrieval. Terms appearing on the list of non-significant words were omitted from the tabulation. In addition, an alphabetized listing of permuted words, referred to among DTIE staff as a "frequency word list," was also prepared. (See Fig. 3.) This listing was quite useful in editing inasmuch as it facilitated correction of spelling errors, identification of additional non-significant words, development of cross-references, etc.

Examination of the preliminary permuted-index printout revealed several problems incidental to inputting and which were not apparent at the time of keypunching. Many titles submitted for keypunching contained hyphenated words and punctuation which caused words followed by no punctuation to be separated from those words that

1	A.E.R.E.
5	A-
2	A-BOMB
1	A-METHOPTERIN
1	ABC
3	ABDOMEN
6	ABDOMINAL
4	ABERRATION
42	ABERRATIONS
1	ABLATION
6	ABNORMAL
8	ABNORMALITIES
1	ABNORMALITY
1	ABOLITION
2	ABORTION
2	ABOVE-GROUND
1	ABROGATION
1	ABSCESS
1	ABSCOPAL
6	ABSENCE
1	ABSORBABILITY

Fig. 3. "Frequency Word List." (The numbers on the left refer to the number of times the word appeared in the preliminary listing of the index.)

```
CRETION OF THALLIUM-204 IN THE RAT. SUGGESTED MAXIMUM PERMISS   1053
OMERULOSCLEROSIS IN THE MOUSE, RAT, AND CHINESE HAMSTER, ASSO  11810
X RAYS AND AQUEOUS DIURESIS IN RAT, RABBIT AND DOG /ATION BY   11798
ON SYNTHESIS OF CITRIC ACID BY RAT- TISSUES IN-VITRO /X-RAYS    5559
EVELS IN STARVED X- IRRADIATED  RATS /OOD AND URINE TYROSINE L 10075
EVELS IN STARVED X- IRRADIATED  RATS /OOD AND URINE TYROSINE L 10078
EVELS IN STARVED X- IRRADIATED  RATS /OOD AND URINE TYROSINE L 10079
OLISM IN X-IRRADIATED MICE AND  RATS /            LIPID METAB 10088
L NICOTINAMIDE IN X-IRRADIATED  RATS /   SYNTHESIS OF N- METHY 10102
LITES IN TOTAL-BODY IRRADIATED  RATS /     INTERMEDIARY METABO 10129
ION DIURESIS BY VASOPRESSIN IN  RATS /HIBITION OF POSTIRRADIAT 10172
TEINURIA INDUCED BY URANIUM IN  RATS /                     PRO 10181
.   / METABOLISM OF ZINC-65 BY  RATS /                         10199
 METABOLISM OF TUNGSTEN-185 BY  RATS /                         10200
AND EXCRETION OF BROMINE-82 BY  RATS /           TISSUE UPTAKE 10228
TIUM IN YOUNG AND ADULT ALBINO  RATS / DIETARY BORON AND STRON 10253
AND STRONTIUM-90 METABOLISM IN  RATS /                CALCIUM  10262
IDIUM- AND CESIUM- SUBSTITUTED  RATS /  ACID EXCRETION IN RUB  10269
AL AND GASTRIC IODIDE PUMPS IN  RATS /                 THYROID 10270
```

Fig. 4. Plural Words Filed Following Singular Words. ("Rat" would be separated from "Rats" by approximately 100 entries if produced by the usual KWIC program. Such words as "Rate" and "Ratio" would therefore precede "Rats" in the normal alphabetical positions.)

had punctuation. Other problems concerned singular and plural forms of a word that resulted in separation of information on the same topic (mouse and mice), variant spellings (haematology and hematology), abbreviations (DNA for desoxyribonucleic acid), element symbols (Fe for iron), and synonymous terms (clotting and coagulation).

Several steps were taken to bring separated terms together in order to improve the usefulness of the index. This was partially accomplished by filing similar words together (see Fig. 4). A list correlating singular and plural forms of words and word variants was established so that words so related could be filed following one another by the computer regardless of their normal alphabetical position.

• Cross-References

A primary deficiency of the usual Key Word in Context index results from the lack of standard language used in the titles that are permuted and causes similar information to be scattered throughout the index. In an attempt to alleviate this problem, an extensive cross-reference listing was prepared using both the frequency word list and the preliminary permuted index. Each word on the word list was studied in relation to the other words. When necessary, the permuted index was consulted to provide some of these relationships. For example, the cross-references under Crustacea, which include Artemia, Astacus, Calanus, Crab, Crayfish, Daphnia, and Shrimp were obtained by review of each permuted word. When it was recognized, for example, that Artemia was a member of the class Crustacea, the term Artemia was listed as a cross-reference under Crustacea. When such word relationships were identified, the word was added to the frequency-word list as a "See also" reference from the more general term. Over 2,000 cross-references were finally identified. The frequency-word list, consisting of about 10,000 words, was easier to scan for word associations and synonyms than reading the more than 60,000 lines of indexing to obtain the same information.

```
ATIONSHIPS OF DNA-BASES OF RAT HAEMATOPOIETIC-TISSUE /LAR REL   8539
, HOMOLOGOUS, AND HETEROLOGOUS  HEMATOPOIETIC TISSUES ON POST-  5745
TREATMENT OF CARCINOMA OF T/    HEMATOPORPHYRIN AND RADIATION.  6410
IS OF HUMAN CANCER / EFFECT OF  HEMATOPORPHYRIN-HG ON GLYCOLYS  2272
OF COBALT-60 GAMMA RAYS ON THE HAEMOCYTES AND CONNECTIVE TISS  11659
                                HAEMOGLOBIN SA METHEMAGLOBIN
                                  OXYHEMOGLOBIN
                                  PORPHYRIN
 / IN-VITRO RADIO- LABILITY OF HAEMOGLOBIN /                    8181
IRRADIATION ON BIOSYNTHESIS OF HAEMOGLOBIN /     EFFECT OF X-   8548
 IRRADI/ QUANTITATIVE STUDY OF HAEMOGLOBIN IN PLASMA OF BLOOD   7846
        / GAMMA RAY EFFECTS ON  HEMOGLOBIN SOLUTIONS /          6683
     / EFFECT OF X-RADIATION ON  HEMOGLOBIN /                   8284
             / REDUCTION OF MET  HEMOGLOBIN BY UV /             7910
ORPTIO/ MODIFIED REACTIVITY OF  HEMOGLOBIN FOLLOWING LIGHT ABS  7958
```

FIG. 5. "See also" References.

```
OSPHORUS-32 BETA-RADIATION  /  HALF-THICKNESS OF DENTIN TO PH   8615
F SCATTERED RADIATION, AND THE HALF-VALUE-LAYER IN A PHANTOM,    929
                                HALIDES SEE EG
                                  BROMIDES
                                  FLUORIDE
        / RADIATION PRESERVED HAM /                             3554
RELATION TO THE COMPOSITION OF HAM /IOACTIVITY OF THE ASH IN    2896
XIDATIVE CHANGES IN IRRADIATED HAMS /                     0     3481
```

FIG. 6. "See" References.

The second computer printout differed from the first in that the correlative words were arranged and printed contiguously. A unique identifying number was also printed with each of the 60,624 lines in the printout. The line numbers identified precisely the position to add cross-references or to delete any superfluous lines permuted to non-significant words.

"See also" and "See" references were inserted in the keyword column as shown in Figs. 5 and 6.

• Publication Format

In order to make maximum use of each line in the index, titles were *wrapped around*. In the final tabulation, the line length was reduced to permit printing two columns per page ready for page makeup.

• Author Index

The author variant problem common to most indexes was minimized in this index by keypunching only the initials of the given names. Transliterations of surnames were reconciled and errors in spelling were corrected. The index was then tabulated in three columns consistent with a preplanned format for use as page copy as shown in Fig. 7.

• Journal Titles

The tabulation of journal citations will be used in-house as a check on the journals scanned by DTIE for *Nuclear Science Abstracts*. This listing of journals will be arranged alphabetically and, within each journal, by volume number. The listing will serve as a frequency count of the radiobiology literature appearing during the three years as covered by the seventeen abstract journals. It will permit selection and subsequent acquisition of those journals having the highest number of relevant articles per year. The listing will also provide a virtually complete tabulation of those journals that publish radiobiology literature.

• Computer Processing

No attempt has been made in this paper to discuss the details of the computer operation. Briefly, all processing was done using the Bell Laboratories Program on an IBM-7090 with the exception that the input for permutation and the output from the IBM were processed on the IBM-1401.

• Conclusions

It is believed that the steps described in this paper have proven sufficiently successful to justify plans to publish another such index for an equally large companion bibliography in the field of radiobiology covering the years 1895-1957.

• Acknowledgment

Acknowledgment is made to Messrs. William M. Vaden and Alden G. Greene, Division of Technical Information Extension, for their helpful suggestions regarding indexing improvements. Sincere appreciation is due Mr. Fred C. Hutton and Mrs. Ada A. Miller, Central Data Processing Division, Union Carbide Corporation, Nuclear Division, Oak Ridge, Tennessee, who were responsible for the entire computer operation.


```
ALLGOTH AM                               2123                           ANGELICO R
   6052                               ANDERSON DL                          5945
ALLISY A                                 8950  10530                    ANGHEL G
   755                                ANDERSON DR                          1765
ALONSOVIVANCOS G                         1615   4960   4961   4962      ANGHEL GH
   8939                                  4963   4964   6718   9943         1657
ALPAUGH EL                              10062  10063  10064  10170      ANGHILERI LJ
   692                                  10171                              11929
ALPEN EL                              ANDERSON EC                       ANGLEMIER AF
   300    301    756   1090              1156   1158   2785   2786         3443   3445
   4855   5066   5684   5685             2787   2788   2903   2904      ANGULO CARPIO MD
   5686   5687   7181   7182             2905   3419  10514                1803
   7183   7903   8439   9268          ANDERSON EP                       ANISIMOVA-ALEXANDROVA VV
   9911  11418  11473  11474            10175  11684                       10869
ALPER T                               ANDERSON JD                       ANKUDINOV VA
   1802   1836   1915   1916             1415                              11797
   5571   5895   6366   6367          ANDERSON JE                       ANLYAN WG
ALPERIN PM                              10426  10427                        4859
   7851                               ANDERSON KP                       ANMA Y
ALPERS JB                                3420                               1680
   9896                               ANDERSON LML                      ANNENKOV BN
```

FIG. 7. Author Index Showing Three-Column Format.

What I am saying, in short, is what I have gleaned from reading science-fiction stories for over thirty years (and nothing I have read or heard from less fantastic sources has changed this opinion) that essentially the problem of useful information retrieval is the problem of setting criteria for usefulness. Once the criteria can be unambiguously established, the rest is a matter of detail. The demon will do your work. All you have to do is tell him, so he can understand, just what characteristics it is that you want him to pass on, and what you want him to reject.

Having said this much, I think in all fairness I should go on to discussing how this sort of thing might some day really be done; not now, but neither as far away as the year 57,000 A.D. I think that there should be real, measurable progress along these lines during our lifetimes—perhaps before we become senior citizens, too.

I can think of at least two ways of handling *broken clock* information, that is, a mass of information which somewhere in its limits contains what we want to have made available to us; and I certainly hope that some of you will take the next step for me, and start working out the engineering specs.

The two ways I have in mind are what I call the "woodlouse" procedure and the "Long John Nebel"[2] procedure.

I don't know how many of you ever have occasion to listen to Long John Nebel on the radio. Long John is on the air some five hours a night, nearly every night. His all-night show's format comprises Long John plus a guest or two—who has some special information that John thinks his listeners will enjoy, plus two or three panelists. I have been one of the panelists two or three times a month for the past five years or so, and I must say that although it has taken large bites of time out of my life, I have enjoyed it. I've met a great many interesting people, and learned a great many things I might not otherwise have encountered. John has some two million insomniac listeners, and among them are representatives of every sort of human being our country can produce—kooks and geniuses, short-order cooks and internes, writers and taxi-drivers, retired ex-presidents and chorus girls. They listen, and they understand, and every once in a while, when someone on the program makes a wrong statement, or can't remember a fact, or leaves out some essential bit of data, they pick up their phones and dial Western Union and send in a telegram containing a correction or amplification.

They function, in short, as a sort of huge pool of information-retrieval specialists.

They are, one might say, the *slow memory* of the *Long John* computer. The *fast memory* comes from the people sitting around the table, before the microphones. Since John is careful to provide a variety of viewpoints, and usually a variety of backgrounds, it is rather probable that most of the data pertinent to a given discussion may well be brought up by one of the men or women in the studio, but, since their storage capacity is finite, it is every once in a while necessary to go to the *slow memory* which has, to all intents and purposes, no upper limit of capacity at all.

[2] His program is broadcast over WNBC, New York.

In other words, what the two-million-unit capacity *slow memory* of the *Long John* computer provides is constant access to immediately available data. It is not necessary to instruct the computer to search its ferrite cores for a given piece of data. The unit volunteers the information without waiting to be asked; it is able to make judgmatical decisions of its own, deciding when a given datum is needed, and then to pass it along.

Of course, sometimes the unit is wrong. Sometimes the datum is either unnecessary or distractive, or just plain fat-headed. But the *fast memory* components sitting around the microphones deal with that eventuality—by rejecting the telegrams that come from misguided souls or from kooks.

What I am suggesting here, of course, is programming along lines which will store information in specialized categories and build into each storage compartment's programming some instruction to let it decide when to volunteer some part of its information. I visualize it as a large collection of sub-units, or sub-routines, in a quite complex computer, each of which is continually scanning all the time a series of regular progress reports on the work and the *thinking* that is going on throughout a university, or a large manufacturing establishment, or a government agency. Most of what passes through each sub-unit is allowed to slip through without emendation or correction or comment. But every once in a while some statement will come through which strikes a chord, so to speak, and then the sub-unit composes a message, addresses it to the proper party and sends it out.

But the critical point is that the sub-unit does not *wait* to be asked. It is a senior faculty coordinator—without decision-making power in the macro scale, of course—and what it does is say, "Look, Charley, you've forgotten about the chi-square law"; or "MacDonald, at Syracuse University, did this same work two years ago with a solid-state laser; I suggest that if you want to do something new you start working with gas lasers."

The other procedure which seems useful to me is that of the woodlouse.

The woodlouse is a small and not very attractive insect which grows, eats, reproduces, dies, and doesn't do much else. Given a choice, it won't do anything else at all. However, every once in a while conditions change in the damp, dark environment of earth and decaying vegetation which is woodlouse heaven, and so it has to do something—or die.

What it does is what is called "klinokinetic behavior." Normally it moves very slowly and sluggishly, and only far enough, in the right direction for reaching something

to eat or perhaps a female woodlouse. When klinokinesis sets in it begins to move quite rapidly. It also changes direction in an essentially patternless, random manner. As long as its environment continues to be uncomfortable, or becomes more so (when it grows too hot, or too dry), it does this on an accelerating scale. As soon as there is a change for the better, it slows down its klinokinetic activity; and when it has thereby blundered once again into a place that's dark enough and damp enough and cheerful enough for it, it goes back to being a sluggish lump.

I have seen one other species exhibit this sort of klinokinesis. I have seen it in writers.

I have in fact seen it in myself, when writing a nonfiction piece or a science-fiction story in which I am trying to make a technical point that I want to have correct. I get up, go to my library shelves and search for a book. If I find what I want, I go back and go on with my typing. If I don't, something quite like klinokinesis sets in; I pick up another book, and another, go to the shelf of back numbers of *Scientific American* or *Scripta Mathematica* or the *Bulletin of the Atomic Scientists*, hunt out my children's *Golden Books of Science* or the back file of *Galaxy*—perhaps not quite as randomly as a woodlouse, but sometimes, I must confess, pretty randomly at that.

Now, I am not a vain man, but I think it is fair for me to say that I am a somewhat more complicated mechanism than a woodlouse.

For that reason, my behavior when I'm hunting for something—I don't quite know what, but I'll recognize it when I hit it—does follow certain rules. It may be that I have a sort of kinesthetic memory which is relevant—I may be pretty sure that whatever it is I'm looking for was somewhere toward the top of a left-hand page, and so I look at a lot of top left-hand pages. I may have a recollection of *when* I saw it, and can perhaps exclude three or four thousand books that I have not opened recently, or didn't then possess. I may remember that it was one particular author, or that it had some connection with some other datum which I can track down in one of these ways.

So that my behavior may not be truly klinokinetic in the sense of being totally random—perhaps certain parameters have been set—but within that area, I hunt until I find it.

Or until I remember it, as sometimes happens.

Or—as happens rather more often than I like to admit—until I decide to give up and go on with something else.

What I am suggestion here is a sort of random scan, of course, coupled with a clear enough statement of recognition factors so that when I or my tame computer ultimately blunder into the desired datum, we will know we have done so.

What all of these things have in common, as you will by now have perceived, is that there is very little attempt at *organizing* information. They are statistical, probabilistic, inherently inefficient mechanisms; they rely on the redundancy of a Long John audience, or the random seeking of a woodlouse.

It is my heartfelt belief that inefficiency of precisely this sort is not a bad thing. I will go farther; it may very well be a good thing. But good or bad, it is certainly a fact of our life in the United States of America at this time.

I think of how our affairs are carried out, and I see a pattern of nearly complete inefficiency in almost all of them. I think of a business office or a laboratory, and I see coffee breaks and conversations around the water cooler and collections for someone's bridal shower. I think of essential public services—firefighting, ambulances—and am aware that where they are volunteer services, as in most suburbs—and probably, for other reasons, even where they are not—the actual putting out of fires and transporting of car-smash victims to a hospital is a sort of minor residual effect left over from the great, and wasteful, effort of putting on fund-raising drives, conducting Firemen's Fairs, keeping an ambulance or a water pump in repair, electing officers and sweeping out the firehouse. If one were to apply vector analysis to the American political structure, one would quickly find that it is the sum of tens of millions of particles, each moving very vigorously in one direction, most of them canceling each other out, which produces a final resultant move of the entire country—minutely—in one direction or another.

We are a rich people.

We have resources enough to afford this sort of profligate consumption of goods and effort. I will not digress far enough to pursue the analogy, although it is a temptation to talk of *directed* societies, and how poorly they are able to use the resources they try so efficiently to plan, but it seems clear to me that efficiency for efficiency's sake is a chimera.

Even if it were necessary—and I do not for one minute think that it is in fact necessary—to use flesh-and-blood human beings for our informational storage units instead of magnetic cores, if it were necessary to use Long John's two million listeners instead of some electronic analogue for them, or real flesh-and-blood men and women to engage in a woodlouse hunt for data instead of a mechanical scan, I still think that it is in that sort of procedure that information retrieval can finally work best and most to the point. After all, we are concerned vitally with the problems of technological unemployment right now. I certainly think there is a paradox in simultaneously fearing out-of-work millions, and at the same time bemoaning inefficient information-handling systems because they might make too heavy a drain on manpower.

But, of course, I think that people need not concern themselves with that sort of thing for very long. I think the machines will do it for us.

All we need is to tell them what to do.

Comments on the 1964 ADI Meeting

A group of attendees and program participants was surveyed after the 1964 meeting of the American Documentation Institute to provide ideas and guidance for future ADI meetings. The responses indicated that attendees view the quality of the technical program as the key to a successful meeting; that a four-day meeting length meets with wide approval; that, if carefully planned, concurrent plenary sessions might be acceptable; that improvements in scheduling are required to permit greater participation in Author Forums; and that conference proceedings greatly enhance the value of the meeting and should be distributed well before the meeting. The many suggestions and criticisms that were offered are being studied in relation to the next regular ADI meeting in 1966.

CARLOS A. CUADRA

System Development Corporation
Santa Monica, California

The 27th Annual Meeting of the American Documentation Institute was held from October 5 to October 8, 1964, in Philadelphia. There were over 760 attendees, almost 10% more than for the previous year's meeting in Chicago. A survey conducted after the 1963 meeting* showed that the meeting was very well received and indicated the general acceptance by attendees of several interesting innovations, in particular the use of preprinted short papers, which were used as the basis for informal small group discussions with the authors. Some of these innovations were retained and even expanded for the Philadelphia meeting.

• Purpose of the Survey

The purpose of the present survey was to obtain additional insights and suggestions that might be helpful in planning the next regular ADI meeting in Los Angeles in 1966.† It was also intended particularly to identify problem areas that should receive attention in planning the technical program, for which the author is responsible.

It should be pointed out that the most valuable guidance for ADI meetings comes out of the experience of those who have planned and conducted previous meetings. These people are intimately aware of the problems and pitfalls that mar the way to a *successful* meeting, and they are aware as well that good meetings depend not only on adequate facilities and planning but also on the quality of the technical reports that the program participants prepare. This survey is a supplement to rather than a substitute for the experience of previous ADI convention planners.

• Nature of the Sample

The present survey was more selective of participants and content than the previous one. Questionnaires were sent to 106 persons, all of whom had attended both the 1963 and 1964 meetings and had signed the 1963 questionnaire. The thirty-nine persons in this group who had conducted Author Forums in 1964 were also asked to comment on their experiences. An additional 25 authors were asked for comments, but were not asked to complete questionnaires. Altogether, 131 participants were contacted, and 66 had responded by the time this report was prepared.

A sample of 66 participants is not large, as surveys go. However, the sample is heavily weighted in the direction of people who take more than a casual role in ADI meetings. Twenty-six of the respondents had conducted Author Forums, and almost half of this group had also been on the 1963 ADI program. Of the 40 other respondents, 17 had been on the 1963 program, and several of these were on the 1964 program as well. Since the primary purpose of the survey was to obtain useful ideas for 1966, there was every reason to believe that suggestions and criticisms from this kind of sample would be worthy of attention.

* C. A. Cuadra, "Evaluation of the 1963 ADI Meeting." *Automation and Scientific Communication, Part III,* American Documentation Institute, Washington, D. C., 1964, pp. 463–471.
† The 1965 meeting of the ADI will be held as part of the FID (International Federation of Documentation) Congress.

• Overall Reaction

Participants were asked: "What was your overall reaction to the meeting? What features and topics were most valuable? Least valuable?" As in the previous survey, those responses which contained a clearly evaluative statement were categorized as Very Good, Good, Fair or Poor. Whereas the preponderance of responses in 1963 were in the Very Good category, the 1964 responses clustered around the Good-Fair portions of the spectrum. A few respondents considered the recent meeting to be excellent. One person commented that the essence of a successful meeting was balance, and that, on the whole, balance was achieved. Other respondents expressed disappointment in the meeting and indicated that it had not been as valuable to them as previous ones.

Most of the individuals questioned cited good and poor features of the meeting. The features mentioned most often as *most valuable* were the Author Forums, the User Information Exchanges, and a specific presentation, that of M. Kessler on the use of MIT's Project MAC for on-line bibliographic search. Also considered valuable features were the Specialist Discussion Seminars and *informal discussions.*

As in the previous survey, the features rated *least valuable* tended to be the nontechnical ones—luncheon, banquet, etc. However, numerous criticisms were also made regarding the technical presentations (papers read at Theme Sessions):

> Several of main sessions not too interesting because of author's presentation . . . and the inability to see most of the slides . . .
>
> Reading of printed papers least valuable.
>
> There should be much more *quality control* on the papers to be presented.

The latter comments indicate—as will the responses to other questions—the importance to attendees of having solid and well-prepared technical presentations. This topic is discussed in more detail later.

• Relative Value of Activities

Participants were asked to rank the various meeting activities which they attended, in order of approximate value to them. The resulting data, shown in Fig. 1, confirm that the technical program, represented by the first five items, is of paramount importance. The other activities and services have limited appeal and limited value, even to those who use them. However, it is interesting to see, from the distribution of value rankings assigned by participants, that the ADI audience is heterogeneous, with interests that probably require diverse activities at their conferences.

As in 1963, the Theme Sessions were the best attended. This year they moved up to rank first in value to participants, in spite of the disappointment expressed by a number of attendees. The Author Forums, which led in 1963, dropped two places, perhaps because the *Proceed-*

Activity	Distribution of Assigned Rankings													Means of Ranked Values	Estimated Attendance*
	1	2	3	4	5	6	7	8	9	10	11	12	13		
Theme Sessions	21	4	9	7	6	4	1		1					2.9	93%
User's Information Exchange Session	7	8	14	3	2	4	2							3.1	70%
Author Forums Based Only on Printed Papers ("Special Contributions")	8	14	6	6	5		3	2						3.2	77%
Author Forums Based on Presented Papers	8	10	11	7	4	4	2	1						3.3	82%
Specialist Discussion Seminars	8	8	8	8	6	3	4							3.5	79%
Exhibits	2	2	5	7	9	8	6	4		1	1			5.3	79%
Informal Reception(s)	5	6	3	4	1	3	6	2	3	2	2	1		5.4	67%
Information Theater			1	2	2	3	3	6	6	2				7.3	44%
ADI Business Meeting	1		2	1	3	4	2	3	2	3	3	3		7.6	47%
Exhibitors' Presentations		1	1		1	2	1	3	1	4	1	2		8.1	30%
Placement Service		1		2	1	1	1		1	2	1	2	3	8.7	26%
Tours				2	1		2		2	3	1		3	8.9	25%
Luncheon/Banquet					1	2	1	1	4	4	2	2	3	9.6	35%

FIG. 1. Attendees' Evaluation of Meeting Activities.

* Figures probably underestimate attendance slightly at "lower-value" activities. Some respondents had obviously attended these activities but did not mark them on the questionnaire.

ings were not available in time to do enough advance reading. The Forums also sustained some additional problems, which will be discussed later.

A comparison was made of the value attributed to the two types of Author Forums (i.e., those based on a preprinted paper and those that, in addition, were presented orally). The results show an even division; that is, as many persons preferred one as the other. While, as we shall see, there is room for improvement of Author Forums, there is no clear basis for choosing whether to have Forums based on papers, talks, or both.

• Length of Meeting

The 1963 ADI Meeting in Chicago was over five days long, and a survey showed that at least half of the participants felt this to be too long. The 1964 meeting was made a day shorter, apparently without detrimental effect. Participants were asked: "What was your feeling about the [4-day] length of the 1964 meeting?" Of the 51 classifiable responses, 62% considered the length good to excellent, and only 6% considered the meeting length to be too short. The other 32% indicated that it may (still) have been too long.

Examination of the individual responses from both years' questionnaires suggests that there are two primary factors that determine the acceptable length of any given meeting. One is sheer time-away-from-work, with or without organizational pressures and limitations on professional meeting attendance. For many people, a five-day meeting is close to or beyond the upper boundary for being away. The second is the apparent quality of the program's content, as witness these responses:

> Seemed to be a little long—perhaps a day shorter and more selective on the material.
>
> Seemed a bit too long for the material that was presented.
>
> For me, this is the wrong question—1 day is too long if papers are substandard. . . . Have we had to pad the program because of lack of papers submitted?

Should the next regular ADI meeting (1966) be made still shorter than four days? Probably not. There is evidence that there was probably sufficient significant material in 1964 to keep a great many participants very interested throughout the entire meeting, but it was not always scheduled conveniently for them. This evidence, taken with the obvious interaction between length and content, suggests that it is preferable not to shorten the meeting. Rather, in view of the apparent compatibility of four-day meetings with the schedules of most ADI meeting attendees, the 1966 meeting should have at least four *full* days, with the expectation of developing a technical program that most of the participants will consider worthy of their time.

• Use of Concurrent Sessions

In every convention there is a good deal of concurrent activity. Even where no major sessions are running simultaneously, there are other activities, such as exhibits or informal contacts, that compete with the formal program. In the 1963 ADI convention in Chicago, the twelve major sessions (Theme Sessions) did not run concurrently. The average participant attended only about eight of the twelve sessions, but in a post-meeting questionnaire, 56% of the participants stated that concurrent Theme Sessions would not have worked well.

The 1964 meeting in Philadelphia involved a great deal of concurrent activity, in one case with 22 Author Forums scheduled to run simultaneously. The Theme Sessions, however, competed only with the exhibits and with informal contacts. Participants were again asked: "From the standpoint of your particular interest, would it have been possible to run several of the Theme Sessions concurrently with each other, in order to leave more time for other activities?" This time, only 25% said "no." Presumably some of the same reasons advanced in the previous survey motivated the "no" responses this year: broad interests of participants and the need for dead time in which to have informal discussions and/or recuperate.

The greater tolerance for concurrent Theme Sessions this year may have been a function of the form of the Sessions. The previous year, each Session involved a panel discussing the state of the art, as reflected in a number of contributed papers. This year, the Theme Sessions consisted almost entirely of delivered papers. Two respondents who said concurrent sessions would have been possible this year specifically qualified this by noting that this would not have been their preference for the previous year.

One person suggested that *theoretical* activities could run concurrently with *applied* activities. This is not an easy division to make, in practice, and it is not clear how well such a division would correspond with the interests of participants. This points up the need to have more specific information on the interest patterns and information needs of the *users* who are served by the ADI meeting. Topics that obviously interest quite different groups are good candidates for concurrent sessions.*

• Author Forums

The high point of the 1963 ADI meeting was the Author Forums, informal meetings in which the author of a contributed paper discusses it (or related topics) with a relatively small group. Participants attended an average of 3.5 such Forums and rated them the most valuable event of the meeting.

The 1964 meeting placed additional emphasis on Author Forums, increasing their number and placing them in prime meeting time. There were seven Author

*We are sensitive to the risk of fostering parochialism. However, scheduling topics of distinctly narrow appeal in plenary sessions has some risks of its own, including loss of the audience.

Forum periods scheduled; thus a person could participate in as many as seven Forums (without Forum-hopping, of course). Participants commonly attended one to five Forums, the average number attended being three—about the same as for the 1963 meeting.

This year's questionnaire asked participants: "Could the effectiveness of Author Forums be improved? How?" Four major suggestions emerged. The first, mentioned in 16% of the responses, was that Forums be scheduled to permit those on the same topics from competing with one another. Examples:

> Better scheduling so that huge numbers are not taking place at the same time.
>
> Instead of having all Author Forums on related subjects meet at the same time, have one from each category meet at the same time. This would allow a person to attend more than one in his field of interest.
>
> Better scheduling . . . I wanted to attend at least three Author Forums at the time I had to run my own.

A second suggestion was that the *Proceedings* be distributed earlier to give people time to do their homework. One person suggested:

> Get papers to advance registrants one to two weeks before meeting, at least as an option (extra for postage paid by recipient).

The third suggestion was that an effort be made to insure that the authors appear, that they start when the schedule says they will, and that cancellations be announced. Examples:

> By making certain that the authors were there (two I wanted to hear did not show).
>
> Some authors did not appear, and nobody knew if they cancelled.

A fourth group of suggestions was for better physical facilities and equipment, including larger rooms, blackboards, and projection equipment.

There were a number of other interesting suggestions, some of which should certainly be considered for the next meeting:

> Program should indicate time and place of Author Forums more clearly.
>
> Perhaps a moderator could be chosen to lead the discussion. . . .
>
> Require each author to present a short (longer if paper was not read or discussed in a Theme Session) explanation of his work first and then have him direct the discussion on the topic in question.
>
> By putting a stop to the invidious heckling.
>
> Awarding Author Forums as a privilege for a good paper, not automatically just for submitting something.
>
> Group the authors for a specific session, so that related questions can be posed to more than one—and the others can join in answering. All too often, one wanted to talk to more than one author.

The authors themselves also provided many suggestions. These will be discussed somewhat later. It should be noted in passing that about 20% of the respondents had no suggestions to make, and others said that the Forums had gone well. As one person put it:

> My own experiences of the Author Forums (or Fora) were happy and showed me the wide variety that is possible—and desirable: author chatting with fellow specialists; author giving a "tutorial" to large audience of people who want to learn; straight technical description and amplification thereof of technical methods/devices, etc.

• Availability Date of Proceedings

The *Proceedings* for the 1964 ADI meeting were distributed at the beginning of the meeting. Those who wanted to participate actively in the Author Forums therefore had to intersperse their scanning and reading with other meeting activities (at least for the first day) or do it late at night. This was also partly the case with the 1963 ADI meeting, for which only a third of the Preprints were available in advance.

Participants were asked: "About how many of the 77 papers in the *Proceedings* did you read before or during the meeting?" The median number of papers read was 7. Three persons read all of the papers (I believe they were reviewers), and about 20% of the respondents indicated that they had not read any. The mean number of Preprints read by participants in the 1963 ADI meeting was 16.2. Since these Preprints were somewhat less than half of the length of the papers in the 1964 *Proceedings*, it appears that the average amount of reading done by convention participants was roughly the same for both years.

Are *Proceedings* necessary at all? To learn how participants use them, this question was asked: "If you had to choose between 1. up-to-the-minute presentations of research findings, with *Proceedings* several months later and 2. less timely findings, but with the *Proceedings* available by meeting time, *which* would you prefer? Why?"

Only 29% of the respondents expressed willingness to accept the *Proceedings* several months later. They stressed timeliness of the research findings:

> Am doing research myself and feel more confident about knowing latest findings.
>
> Isn't the purpose of the meeting to get oneself updated?
>
> No one reads them anyway; and besides, why mull over old stuff?

> *Proceedings* take time to read and digest. Meanwhile the author is not asked the proper questions about his outdated work. He may as well talk about his latest results.

Almost two-thirds of the responses rejected the idea of waiting several months for *Proceedings*, noting that having the *Proceedings* available is necessary to get maximum benefit from the meeting:

> . . . gives a basis for deciding which papers are most useful to the individual attendee as well as the questions he would like to ask.
>
> The meeting period is the best time to read and clarify with the author the contents of the papers.
>
> Post-meeting *Proceedings* are largely unread. . . . this gives the reference available at the meeting to check points in the presentation.
>
> One likes to be prepared to discuss papers presented. I realize that a great many people can discuss almost anything on no notice, but many cannot.

The majority also argued that it was not necessary to use the *Proceedings* for up-to-the-minute reports:

> I don't think a month or two makes that much difference in up-to-dateness.
>
> Don't think that our field advances so rapidly that several months would make a difference.
>
> Up-to-the-minute research findings I can get from informal discussion in narrow field of my own interest. All else is new to me even if a year old.
>
> The small group needing such up-to-the-minute data know and can correspond with one another. . . .

Several persons, even including one who was willing to accept "late" *Proceedings,* pointed out that having the *Proceedings* available by meeting time can improve the technical quality of the papers:

> I think that it makes for a better organization of material presented at the meetings.
>
> It affords the program committee more time for review and selection.
>
> . . . it tends to insure that those who wish to present papers at the meeting will actually have something to say, rather than merely agitate the air, as in the good old days of ADI.

Several persons questioned whether *Proceedings* was a misnomer, arguing that most of the public presentations are *ad lib.* It was suggested that having short Preprints first and then *real Proceedings* later (as was done in 1963) would be the most helpful.

It is evident that timeliness in the reporting of the research findings is not without merit. However, it is also clear that most participants do not want to purchase that timeliness at the cost of delaying the publication of materials which are of value in the actual participation in the meeting. There is no doubt whatever that the value is enhanced if the material—whether it be called Preprints or Proceedings—is available to the participants in sufficient time for them to read selected parts and plan their meeting participation. As several participants suggested, it is not out of the question, with modern publication techniques, to have such material timely in both senses in which we have been discussing it.

• Effect of Proceedings on Meeting Participation

Some speakers are hesitant to read a paper they know the audience may have already read, and they either talk *around* their paper, mystifying those who have not read it, or they deliver a more time-consuming and less organized account of the same thoughts contained in their paper. In such instances, the availability of preprinted Proceedings may actually reduce the effectiveness of the presentation.

To determine whether this phenomenon occurred in the 1964 ADI convention, participants were asked: "How would your own participation in the Meeting have been different if no *Proceedings* were available until after the meeting?" Thirty-six percent of the response indicated that it would not have made any difference at all, and sixty-four per cent indicated that there would have been some difference. However, no one stated or implied that *Proceedings* have any deleterious effect upon author-audience-communication.

One person who said "no difference" noted that the *Proceedings* were "already out of date." Two others indicated that they did not need the *Proceedings* to plan the use of their time at the meeting. Another person who said "no difference" pointed out that this would not necessarily have been true for the 1963 meeting:

> In 1963, when panels discussed the printed contributions, it was possible to refer to particular papers and decide on Author Forums (to attend).

Those who felt that without the *Proceedings* they would have participated differently gave several reasons. About a third of these responses suggested that the *Proceedings* help one to understand and participate more intelligently in Author Forums and discussion:

> Would probably have meant less-informed discussion at my Author Forum.
>
> The Author Forums would be less valuable.

Another group indicated that without the *Proceedings* they would have attended more of the sessions or would have been more active listeners. It is significant that *no one* considered either of these a virtue:

> I would probably have worked harder on "listening," but what would that have accomplished? . . .

I might then have attended certain presentations for which the *Proceedings* served as fair warning.

I would have been forced to sit through a lot of dull papers to be sure that I didn't miss something of value.

A third group said that without the *Proceedings,* they would not have been able to plan the use of time as effectively:

I would not have known which sessions to attend. Without *Proceedings,* I would be dependent.

I might not have attended the terrific presentation by the boys from M.I.T. . . .

Two respondents who had run Author Forums noted that without the *Proceedings* they would have been compelled to present a formal paper.

These responses, together with those previously discussed, indicate the strong reliance that ADI meeting attendees place on some form of *Proceedings* to guide their choices of activity and enhance their level of participation. It is obvious that *Proceedings* have an important role to play in ADI meetings and that a significant amount of attention should be given to integrating them effectively with the overall technical program.

• Suggestions for Specific Topics and Speakers

Participants were asked to indicate topics they would like to see emphasized, as well as individuals whose participation would enhance the technical quality of the next meeting. At least 50 reasonably distinct topic suggestions were made, a few of them being seconded by a number of individuals. For example, more than a dozen responses called for more information on demonstrations of actually operating, *practical* systems. Several of these responses expressed annoyance at the plethora of proposed systems and techniques and expressed a strong interest in real experience. One person asked pointedly for "some honest reports on the operating characteristics of the government systems."

The second most popular topic concerned the user: studies of user's needs, the methods used in such studies, and user-oriented systems. One person called for "a philosophical discussion of the purpose of user studies."

Also mentioned several times were education and training, library mechanization, associative memory techniques, and the behavioral (as opposed to technical) aspects of communication. Some unusual topic suggestions that might prove interesting include the following:

Costs of *R&D* in information science from standpoint of cost-effectiveness.

An appraisal of the purpose, structure and performance of ADI.

Criteria for evaluating the performance of designers and managers of developmental and operating systems.

Scientific methodology in documentation research.

A review and evaluation of what various disciplines have contributed to documentation, e.g., statistics, linquistics, topology, programming.

Analysis of techniques for . . . system design.

Examining where we've come during the past 20 years—what are the patterns?

National planning.

May I suggest that much effort be given to spotting "young talent" at technical colleges, universities, etc., and encouraging them to contribute in 1966?

Several persons who did not make suggestions of specific topics observed that ADI convention planning already seemed to be pretty good, and one passed the buck right back: "Use your judgment."

Twenty-nine individuals were specifically suggested as people whose presence and participation would enhance the next program. Several of them are among the *Who's Who* in documentation that did not play a very active role in the 1964 meeting. Others are known to have been participants in the meeting running almost concurrently at Pittsburgh. (As one person put it, "How come that *¢@#! thing in Pittsburgh was scheduled the same time as our meeting?") The planners of the next ADI technical program will certainly want to consider encouraging many of the suggested individuals to take an active part in that program.

• Suggestions for Improving Future Meetings

The final question asked was: "What general suggestions, not otherwise given above, do you have for improving future ADI meetings?" About three-fourths of those who returned the questionnaires provided suggestions, and fully a third of these responses called for means for improving the quality of the technical presentations. This was true both for respondents who had been on the program and for those who had not. Here are some of the responses:

More screening of presented papers so only best are presented verbally.

. . . the greatest service an ADI meeting can perform is to guarantee *quality* in the papers presented, even if meeting should be held every two years instead of every year. . . .

Get rid of the great number of poor-quality papers (ADI could get discriminating reviewers, if it were willing to use them).

Stricter guidelines for presentation, performance and slides should be adhered to.

Firm control by session chairmen as well as by general management of the quality and quantity of materials submitted for publication and for oral presentation.

The second-largest group of suggestions had to do with schedules:

> I frequently found conflict of interest in the special contributions. Somehow these should be worked out on a sequential time basis rather than all running at the same time.
>
> Too many simultaneous Author Forums in the same field so that it is impossible to attend several of them. Concurrent tutorial and research-level sessions, each limited to 1½ hour, and not scheduled concurrently with tours or demonstrations.
>
> The schedule on the user exchange sessions was poor. Some of each type of session each evening would have better satisfied me.
>
> Please schedule one afternoon or morning for tours ONLY so an attendee can participate in one or two without missing sessions and/or seminars, etc.

A number of persons suggested changes in convention strategy or content, for example:

> Return to review-panel presentation of Theme Sessions, and thereby discard "Author Forums for papers not presented."
>
> Cut the list of convention activities to about half which are presented in a straightforward schedule that people can follow.
>
> [Plan on] . . . three levels—1. frankly tutorial and *advertised as such*, 2. general interest material, 3. specialist seminars *advertised as such*.
>
> Short courses in IR for those who are new in the field. A glossary of terms in the program since speakers frequently use *a.k.* IR terms not generally understood.
>
> Each meeting have a session devoted to the communication needs of a particular discipline Include both subject experts . . . and documentalists. . . .
>
> Ensure that the subject area is specific enough that discussions can avoid the inevitable generalities about documentation in all of science.
>
> More demonstrations and manufacturer exhibits.

Several suggestions were received this year on the need for greater courtesy to speakers:

> I think ADI membership lacks a great deal of just plain courtesy. I belong to quite a number of organizations, but I do not believe I have ever seen a more rude group or one which lacks the basics of human kindness.
>
> I realize it would be difficult to prevent bickering and the barbed comment but why display it in public as we did in the first session. . . .
>
> While the criticisms of the panels at Chicago may have been severe, at least they were 1. heard, 2. seldom rude.

Several people offered suggestions regarding facilities. They included providing maps of the meeting facilities, supplying spring water, avoiding hotels with noisy air conditioning, and making it possible for all members and attendees to get into the convention hotel. Comments were also made on the need to get convention materials out sooner. One person even offered a schedule:

> The programs, announcements, et al, should be available at least 60 days prior to the meeting. *Proceedings* or Preprints should be in the hands of registrants by mail at least 10 days prior to the meeting. Preprints of speeches or papers not in the *Proceedings* should be available in quantity at the time of the meeting.

A final group of miscellaneous comments included issuing the registration list by noon of the second day, restoring first or given name, and distinguishing Miss or Mrs.; having more receptions and other opportunities for informal communication; providing a listing of "who knows what?" so that members could approach *authorities* in various areas to get informal tutoring and consultation; announcing administrative meetings of ADI personnel ahead of time and holding them at some time that does not conflict with other commitments; reviving the *documentalists bookshelf*, which was tried for a couple of years and was apparently inquired about this year; making coffee available for purchase at breaks; and handing out this kind of survey at the meeting "while people's woes are fresh in their minds."

• Special Comments From Authors

A number of authors were asked, apart from (or instead of) the questionnaire to comment on their Author Forums experience, as well as on other aspects of the meeting. The responses were extremely thoughtful, and it is impossible to do justice to them in a short space. Perhaps the best that can be done is to present a composite letter, representing the prevailing thoughts and, wherever possible, the actual words of the authors' letters:

Dear Dr. Cuadra:

I would like to comment a little more than space allowed in your questionnaire.

First, with respect to my Author Forum, I feel that it went quite well. I was pleased with the turnout. I first ascertained that most members of the audience had not had an opportunity to study my paper, so I gave a brief summary. I think that they expected me to run over it again, possibly with additional information or with a slightly different point of view from the published version.

The *Proceedings* were a great help for the Forums, particularly in allowing those who had read them to *break the ice*. Once this was done, the discussion proceeded well. We arrived at no earth-shaking conclusions, but we did spend a stimulating hour or so together. Members of the audience were interested and I believe they had a good time.

I am sure that Author Forums can be made more effective and less trying for the authors. I could have used a blackboard and perhaps other equipment. Also, because the main meeting was running late, I started out with a smaller group and then had to tax my ingenuity to cope with the newcomers that came in much later. While I'm talking about schedules, I should probably add that there ought to be some way to avoid having so many of the Author Forums in a particular subject area overlap. This causes much conflict and session-hopping, the latter a difficult task with rooms spread out from basement to fifth floor.

It would also seem desirable to try to level out the tremendous variation in activity that goes on at the author session, varying from zero to very good. One idea might be to give the authors a better orientation on what is expected (I never even met my session chairman). Perhaps it might be desirable to have written questions directed to the authors in advance to help them organize their Forums better. Another idea might be to have a presiding officer or moderator for each Forum, to help get things started and to help cut off utterly extraneous discussion. Still another possibility is to have a panel of several authors who had presented short papers in a given subject matter.

One thing I did for my Forum may have some wider utility: I handed out printed copies of my paper at the beginning of the Forum, and I feel this added significantly to the audience's ability to participate in the discussion that followed. This might conceivably be considered as an alternative to the bound *Proceedings*. Some of the papers in the *Proceedings* (and mine is one of them) don't deserve that kind of permanence.

I personally believe that the Author Forum idea has very great potential, because the greatest value of meetings lies in the opportunity for face-to-face discussion. To develop this potential fully is going to require a lot of intensive work on the part of those connected with the meeting. We must be particularly careful not to place so much emphasis on informal activities that we completely minimize the importance of preparing and delivering solid technical papers.

With respect to the other aspects of the meeting, I should say first of all that it appeared to be very well organized. Except for some missing blackboards and projection equipment not available on time, the facilities were good. I am sure this reflects careful attention to detail on the part of the hosts.

The Theme Sessions seemed aimed too strongly at conceptual and theoretical work. They could have been improved by more discussion of operating systems. While we naturally cannot emphasize only talks dealing with how-to-do-it on operating systems, the papers should be better tuned to audience interests and background. Particularly, I believe that papers chosen for plenary sessions need to have broadness of view or they cannot hold the attention of the extremely diverse group that attends ADI meetings.

I have several other specific suggestions that you may want to consider for the next meeting:

1. Make sure that there is good communication with participants before the next meeting. Acknowledge receipt of their abstracts and/or papers, and notify them (early(if they have or have not been accepted.
2. Get the announcements, program, registration materials and Preprints or *Proceedings* on the meeting out much earlier. To be useful, the *Proceedings* should be out at least two weeks ahead of time.
3. Issue specifications for slides and length of papers to avoid some of the unfortunate presentations that were made.
4. Maintain meeting control and adherence to the schedule.

I hope what I have expressed will be some help to you in your very difficult task as 1966 Technical Sessions Chairman. Good luck.

(signed) A Composite

• What Now?

The primary purpose of this survey was to obtain insights and suggestions to help in planning subsequent ADI meetings beginning with the one in October 1966. It has certainly done that, not simply by providing ayes and nays to specific questions—one is never clever enough to ask all the right questions—but by illuminating the requirements that underlie each suggestion, compliment, or complaint. It is these implicit user requirements that convention planners must translate into specifications for the 1966 ADI meeting. The survey may also provide additional perspective to ADI members on the diverse interests, needs, and attitudes of their fellow members. To the extent that effective presentations require knowledge of the audience, the survey may contribute to better ADI presentations.

The planning work for 1966 has already begun. The many suggestions offered are being considered and weighed against the experience from past meetings and the preliminary concepts for the next meeting. Over time, the ADI has evolved an excellent formula for meeting the very diverse interests of its members, and no need for radical change is evident. The ideas and suggestions contributed by the survey participants will help to insure that the evolution continues in the direction of greater service to ADI members.

Publication Patterns of Scientific Serials*

American libraries are undertaking computer-controlled inventories of serials holdings. Such automated records are posited on predictability of serials receipts. Publication patterns of serials, falling into 24 distinct classes and two further categories, are developed for machine-based prediction and record updating. The relationship between publication patterns and frequency/dating systems is examined. The publication patterns listed are considered universally applicable to machine methods.

DAVID BISHOP, ARNOLD L. MILNER,
AND FRED W. ROPER

Biomedical Library, Center for the Health Sciences
University of California, Los Angeles

The overriding importance of serial† publications in the sciences, and increasingly in other disciplines, is an axiom of contemporary documentation. Conventional methods of recording and gaining access to the serials holdings of libraries have proved inadequate in the face of their continuing importance and growth. In an attempt to gain more efficient control of their serials holdings, a number of North American libraries have turned to computers to provide current inventories which give total and accurate listings of available serials holdings. Such computer-produced listings have the added advantage of being capable of arrangement in many ways as well as of distribution to the library's clientele outside the library.

The translation of conventional serials records into a form convenient for use with the new technology has necessitated determining a previously unneeded, and hence unstudied, phenomenon: the publication patterns of serials. The growth of serial publications through the centuries has been at best haphazard, at worst anarchic, and the most optimistic view of their future development indicates no change in this condition. Particularly haphazard has been the method whereby the serial publisher has identified the successive issues of his publication.

In fact, two independent but interacting identifying systems have been used side by side, in most cases both systems being applied to any one serial. One of these systems is, as it were, an internal numbering system, whereby successive issues are given unique identities by some sequential numeric values. (These are the familiar volume, volume-and-issue, volume-and-part, issue, Band, Band-Heft, etc. values, assigned in either continuous or repeating series.) The other is a chronological, or dating, system whereby successive issues are identified by their nominal date of publication.

Some form of the chronological system is always present, though not always sufficient for unique identification. The logical criteria *necessary and sufficient* are used to determine the identifying system characteristics of any one serial. An internal numbering system is usually present and characteristic. When it is absent, the chronological system is sufficiently precise for unique identification and can then be considered a type of numbering system. It is the internal number system (or in its absence the specific chronological system) which we have termed the *publication pattern.*

This paper will discuss publication patterns in detail and includes discussion of the related concepts of frequency and interval. For clarity, these topics may be outlined as follows:

Publication pattern (internal numbering system):

	(a)	volume	issue	year
or	(b)	volume	....	year
or	(c)	......	issue	year
or	(d)	......	....	year

Frequency: periodic division(s) of year(s) per publication unit. (Monthly, quarterly, etc.)

* The work on which this study is based was supported in part by a General Research Support Grant from the United States Public Health Service to the School of Medicine, UCLA.

† For the purposes of this paper, the following definition of a serial is used: "Publications of indefinite duration appearing in sequence (regularly or irregularly) under a common title, their order being ascertainable from numbers or dates appearing in each issue." Int. Assoc. Agric. Libns. & Docs. *Qtly. Bull.* 9(3): 180, July, 1964.

Interval (an artificial numbering device):
1st, 2nd, 3rd . . . *n*th publication unit per year(s).

Practically all patterns are *modified*, one is tempted to say, restrained, by an old concept: frequency of publication. It cannot be overstressed that the concept of publication pattern is distinct from that of frequency of publication. Thus it is quite possible, and in practice quite frequent, for a serial to be issued monthly, bimonthly, quarterly, etc., and have the same *pattern* of numerical identification. There is, however, a subtle connection: the vast majority of serial publications fall into the sub-class *periodicals* (sub-annually published serials with more or less fixed publication dates equally divided within a year). The very periodicity of issuance of such publications originally suggested identifying numerical patterns, the most common and best known being to make all issues within one year equal to one *volume*, with each periodical issue within the year being given a sequential *issue* number.

In fact, some of the publication patterns enumerated below are essentially a statement of frequency (e.g., the date-identified annual publication). Most have a built-in frequency limitation reflected in their pattern of repeated numbering; some few discard the concept completely. In *all* cases the concept of publication pattern is on a higher generic level and takes into account both frequency-related patterns and patterns of non-periodical serials.

Where a definite frequency is used by the publication, the frequency division can be given a numeric value and stand in place of the more conventional months, seasons, quarters, etc. We call such year divisions *intervals*. The interval, however, is a broader concept than frequency. Where the latter is properly confined to periodic divisions of years, the interval device includes also aperiodic divisions. For example, a journal publishing four issues a year at regular periods has a quarterly frequency, with intervals numbered 1 through 4. One publishing four issues per year, but all four, say, in any one month, has an irregular frequency, but these issues may still be given interval values 1 through 4.

In this way the first issue of a new year of any journal can be designated *interval 1*; its issuance date may be Jan., Winter, Jan.-Mar. or what-have-you; its publication pattern identification may be any value. In practice, with the large majority of periodicals having repeating issue-number cycles for each calendar year, the issue number and the interval may be identical (e.g., of a monthly journal, the vol. 20, no. 10, October issue will also be *interval 10*). The interval is nonetheless independent of the publication pattern: in the case of repeating issue-number cycles (by six-month periods), an issue of another monthly journal may be vol. 20, no. 3, October, but the interval will still be *10*.

This device obviates the need for a built-in computer calendar for predicting the publication date of the next expected issue. Merely, the interval is updated by *1* for each issue until the known total number of issues for the year is reached. The interval for the next expected issue is then re-set at *1*, and the *year* of the next expected issue is updated, regardless of its numbering within its publication pattern. (For annuals, biennials, etc., the interval is always *1*, and the updating is performed by checking frequency and adding the necessary increment to the year.) The inclusion of the interval on the check-in card for the next expected issue affords an added safeguard for check-in, being in effect a date indicator to accompany the publication-pattern numbering system.

The old concept of frequency, therefore, is still of importance as reflected in the *interval* device (and is also a necessary component of date-predictive perpetual-calendar systems). It is, nonetheless, secondary to the new concept of publication pattern which is essential for prediction of the bibliographically identifiable (numbered) publication unit.

The publication patterns that we have determined fall into 24 distinct classes, with a further two classes that require some verbal hedging. Our patterns have been drawn from studies of the more than 10,000 serials in the collections of the UCLA Biomedical Library where a computer-based serials record is currently under development.* This special collection, limited to the biological and health sciences in their widest sense, may not provide examples of all serial publication patterns. Discussions with librarians working closely with serials in other disciplines, however, have failed to bring further patterns to our attention. The present schema, therefore, is presented as a tentative but hopefully useful first compilation of this necessary information.

It would have been possible to list the 26 classes perceived in a number of arrangements, the most likely-seeming being that of frequency of use. For their usefulness for machine-based systems we have instead presented them in a logical sequence whereby their controlling characteristics can be used as a signalling device. Actually, we have departed from strict machine logic in order to have the most frequently encountered pattern first; logically it would have been in second place.

Logis Breakdown of Serial Publication Patterns

	Patterns
A. Serial has a *volume number*	
1. The issue (volume subdivision) is a *number*	1-5
2. The issue is *blank* (no volume) subdivisions)	6-7
3. The issue is designated by *letters*,* i.e., the issues are designated by *seasons*	8
or the issues are designated by *months*	9

*The system, it is stressed, is not yet operational. This paper is based on design studies only.

TABLE 1

	VOLUME		ISSUE			YEAR					
	1 1	1 1	1 2	1 2	1 2	1960	1960	1960	[1959]	Unpredictable. Varies from many volumes per year to many years per volume.	Unpredictable. Varies from many issues per year to many years per issue.
	1 1	1 1	3 4	3 4	3 4	1961			1960		
	2 2	2 2	5 6	5 6	1 2	1962	1961				
	2 2	2 2	7 8	7 8	3 4	1963			1961		
	3 3	1 1	9 10	1 2	1 2	1964	1962	1961			
	3 3	1 1	11 12	3 4	3 4	1965			1962		
	4 4	2 2	13 14	5 6	1 2	1966	1963				
	4 4	2 2	15 16	7 8	3 4	1967			[1963]		
Pattern 1	X				X	X	X	X	X		
Pattern 2	X		X			X	X	X	X		
Pattern 3	X			X		X	X	X			
Pattern 4		X	X					X			
Pattern 5		X			X			X			
Pattern 7	X									X	
Pattern 10					X		X				
Pattern 11			X				X				
Pattern 12			X								X

TABLE 2

	YEAR						
	1960	1960/1	1960/1	1960	1960	1960	1960
	1961	1961/2	1962/3	1962	1963	1964	1965
	1962	1962/3	1964/5	1964	1966	1968	1970
	Annual (calender)	Annual (fiscal)	Biennial (covers 2 years)	Biennial (issued every 2 years)	Triennial	Quadrennial	Quinquennial
Pattern 13	X						
Pattern 14		X					
Pattern 15			X				
Pattern 16				X			
Pattern 17					X		
Pattern 18						X	
Pattern 19							X

B. Serial does *not* have a *volume number*

1. The issue (publication unit) is a *number*	10-12
2. The issue is *blank* (i.e., there is neither volume nor issue numbering, but only chronological identification)	13-19
3. The issue is designated by *letters*,* i.e., the issues are designated by *seasons*	20
or the issues are designated by *months*	21-24
C. Unpredictable	25-26

From the above and the accompanying tables, it is now possible to identify publication patterns 1 through 24. For clarity, however, a brief description of each is given. It should be noted that the use of four issues per volume in the tables is for convenience only and implies no limitation in practice.

PATTERN 1:

Volume and issue numbering present; fixed number of issues per volume; fixed number of volumes per time period. Volume numbering continuous. Issue numbering continuous within each volume, restarting at *1* with each new volume. May be 1 volume per year, 2 or more volumes per year, 1 volume for 2 or more years, or volume may not coincide with calendar year although published in a period based on calendar year. This latter occurrence may best be defined by an example: thus volume 1 of a given title may run from July of one year through June of the next; volume 2 will then run from the following July through the following June, etc. This is sometimes, though not always, concurrent with a normal fiscal year. Henceforth we shall call such occurrences by the inelegant but descriptive term *flop-over*.

PATTERN 2:

Volume and issue numbering present; fixed number of issues per volume; fixed number of volumes per time period. Volume numbering and issue numbering both continuous, issue numbering never re-starting at *1*. May be 1 volume per year, 2 or more volumes per year, 1 volume per 2 or more years, or *flop-over*.

PATTERN 3:

Volume and issue numbering present; fixed number of issues per volume; fixed number of volumes per time period. Volume numering continuous. Issue numbering continuous through a set number of two or more volumes, then re-starts as *1* to repeat cycle. May be 1 volume per year, 2 or more volumes per year, or 1 volume for 2 or more years.

* Letter-designation of an issue is disregarded where *both* an issue number and a month/season issue designation occur. That is, an issue number, when present, determines the publication pattern whether or not the issue has also an identifying sub-annual date.

PATTERN 4:

Volume and issue numbering present; fixed number of issues per volume; fixed number of volumes per time period. Volume numbering coincident with half-year: i.e., volumes are numbered *1* for each Jan-Jun period and *2* for each Jul-Dec period, repeating this cycle yearly. Issue numbering is continuous. Two volumes per year only.

PATTERN 5:

Volume and issue numbering present; fixed number of issues per volume; fixed number of volumes per time period. Volume numbering the same as in PATTERN 4: i.e., vol. 1 designates Jan-Jun of any one year, vol. 2 designates Jul-Dec of the same year. Issue numbering continuous within each volume, re-starting at *1* with each new volume. Two volumes per year only.

PATTERN 6:

Volume number, but no issue number. Volume may be annual, biennial, triennial, etc., or *flop-over*.

PATTERN 7:

Volume number but no issue number. Irregularly issued: varies from many volumes per year through many years per volume. Only the next volume number is predictable.

PATTERN 8:

Volume numbering and issue designation by season; fixed number (4) of issues per volume; fixed number of volumes per time period. Issue designation repeats annually. Each volume covers one year only, but may not be coincident with calendar year (i.e., may be *flop-over*).

PATTERN 9:

Volume number and issue designation by month; 12 issues per volume; fixed number of volumes per time period. Issue designation repeats annually. Each volume covers one year only, but may not be coincident with calendar year (i.e., may be *flop-over*). Note that in the accompanying Table one example only is given of *flop-over:* a volume starting in July of one year and ending in June of the next. In fact, a volume may start at *any* point in the year (as in Table 4—Seasons) and proceed for a 12-month period.

PATTERN 10:

No volume numbering, but issue numbering; fixed number of issues per time period. (Differs from PATTERN 6 in that publication unit is sub-annual.) Issue numbering is repeating: i.e., continuous numbering for any given year, re-starting at *1* for each succeeding year.

PATTERN 11:

No volume numbering; but sub-annual issue numbering. Issue numbering is continuous, with fixed number of issues per year.

PATTERN 6

TABLE 3

VOLUME	ISSUE	YEAR					
1	-	1960	1960/1	1960	1960	1960	
2	-	1961	1962/3	1963	1964	1965	1960
3	-	1962	1964/5	1966	1968	1970	1961

PATTERN 8

TABLE 4

VOLUME	ISSUE	VOLUME	ISSUE	VOLUME	ISSUE	VOLUME	ISSUE	YEAR
1	WINT							
1	SPRG	1	WINT					1960
1	SUMR	1	SPRG	1	WINT			
1	FALL	1	SUMR	1	SPRG	1	WINT	
2	WINT	1	FALL	1	SUMR	1	SPRG	
2	SPRG	2	WINT	1	FALL	1	SUMR	1961
2	SUMR	2	SPRG	2	WINT	1	FALL	
2	FALL	2	SUMR	2	SPRG	2	WINT	
3	WINT	2	FALL	2	SUMR	2	SPRG	
3	SPRG	3	WINT	2	FALL	2	SUMR	1962
3	SUMR	3	SPRG	3	WINT	2	FALL	
3	FALL	3	SUMR	3	SPRG	3	WINT	
		3	FALL	3	SUMR	3	SPRG	
				3	FALL	3	SUMR	
						3	FALL	
		Flopover		Flopover		Flopover		

PATTERN 12:

Same as PATTERN 11, except no fixed number of issues per year, so that the only predictable value is the number of the next issue. Varies from many issues per year to many years per issue. Differs from PATTERN 7 only in the publisher's designation of the publication unit as *volume* or *number* (or their foreign equivalents).

PATTERN 13:

No volume or issue numbering. Publication is identified only by year of publication, i.e., by its annual (calendar) frequency (1961, 1962).

PATTERN 14:

Same as PATTERN 13, but identified by its fiscal-year frequency (1961/2, 1962/3).

PATTERN 15:

Same as PATTERN 13, but identified by its biennial coverage (e.g., 1962/63, 1964/65).

PATTERN 16:

Same as PATTERN 13, but identified by its biennial issuance. Differs from PATTERN 15 in not indicating biennial coverage (e.g., 1962, 1964, 1966, etc.).

PATTERNS 17-19:

Same as PATTERN 13, but identified by triennial, quadrennial and quinquennial issuance dates respectively. (These classes comprise so few titles, that no distinction is made between coverage dates and issuance dates as was done for PATTERNS 15 and 16.)

PATTERN 20:

No volume numbering, issue designated by season. Four issues per year, no cases known other than those concurrent with calendar year. Issue designation repeats annually.

PATTERN 21:

No volume numbering, issue designated by month. Twelve issues per year, concurrent with calendar year. Issue designation repeats annually.

PATTERNS 22-24:

Same as PATTERN 21 except that issues are published every two, every four, or every six months.

PATTERN 25:

This pattern is a special case of PATTERN 1, and relates to serials which have continuous volume numbering, issue numbering repeating for each volume, and a fixed number of issues per volume. However, there is not necessarily a fixed number of volumes per year, and the issues of a number of volumes are published concurrently. That is to say, while volume numbers and issue numbers are predictable, frequency of issue is highly erratic so that one cannot predict what volume number may be on the next issue received; a number of volumes are issued concurrently but not coincidentally. That the issues within each volume are predictable permits a measure of control for the next-expected-issue, but special programming is necessary for computer handling of such serials. Unfortunately, though not a common pattern, this pattern is found in a number of extremely important biomedical titles. Its method of control will depend on the automatic serials record system used.

PATTERN 26:

Were it not unfashionable to do so, this pattern would bear the designation *Miscellaneous*. By definition it covers all publication patterns not included in PATTERNS 1-25 above. In practice it has worked out to the following sub-classes:

(a) Sporadic issues only are purchased by the library. This is particularly liable to happen in a subject-oriented library. While the selected issues may well be numbered in one of the regular patterns, it is not possible to tell in advance which issue or volume will be purchased.

(b) A widely-fluctuating number of issues per volume in a PATTERN 1 type of serial. While a computer program can be written to handle *over-age* and *under-age* issues, in practice this must have some limitations. (In our own system we have limited it to two issues over or under the expected number.)

(c) Any case where the pattern is definite, but peculiar to that serial or to so few serials that it was not felt worthwhile to specify it in PATTERNS 1-25. Such patterns would require special programming unless handled in the "Unpredictable" class, and the economics of such action become a major factor. (As an example, the UCLA Biomedical Library has a serial issued with numbers 201 through 212 each year, but with number 207 never published. Easily programmable, but is such special treatment warranted? We think not.)

• Example of Use of Publication Pattern

The important British medical journal *Lancet* is an example of a PATTERN 4 serial, having both volume and issue numbering with a fixed number of issues per volume and a fixed number of volumes per time period. The volume cycle is repeating: volumes are numbered *1* and *2* each year, the first indicating Jan-Jun, the second indicating Jul-Dec. The issue numbering is continuous.

Specific information for this title stored in the computer (other than the identifying information for PATTERN 4) is the number of issues per volume (26), whether or not the volumes are *flop-overs* (note: PATTERN 4 has no such possibility in any case) and the number of issues per year (52).

Assume the base input started with the first issue of 1964, i.e., vol. 1, no. 7323, 1964, interval 1. This issue number is stored in memory, being the first issue of the

TABLE 5

	VOLUME		ISSUE					YEAR	
	1		JAN.	JAN.	JAN.	JAN.	WINT		
	1		FEB.						
	1		MAR.	MAR.					
	1		APR.		APR.		SPRG		
	1		MAY.	MAY.					
	1		JUN.						
	1	1	JUL.	JUL.	JUL.	JUL.	SUMR	1960	
	1	1	AUG.						
	1	1	SEP.	SEP.					
	1	1	OCT.		OCT.		FALL		
	1	1	NOV.	NOV.					
	1	1	DEC.						1960
	2	1	JAN.	JAN.	JAN.	JAN.	WINT		
	2	1	FEB.						
	2	1	MAR.	MAR.					
	2	1	APR.		APR.		SPRG		
	2	1	MAY.	MAY.					
	2	1	JUN.						
	2	2	JUL.	JUL.	JUL.	JUL.	SUMR	1961	
	2	2	AUG.						
	2	2	SEP.	SEP.					
	2	2	OCT.		OCT.		FALL		
	2	2	NOV.	NOV.					
	2	2	DEC.						1961
	3	2	JAN.	JAN.	JAN.	JAN.	WINT		
	3	2	FEB.						
	3	2	MAR.	MAR.					
	3	2	APR.		APR.		SPRG		
	3	2	MAY.	MAY.					
	3	2	JUN.						
	3	3	JUL.	JUL.	JUL.	JUL.	SUMR	1962	
	3	3	AUG.						
	3	3	SEP.	SEP.					
	3	3	OCT.		OCT.		FALL		
	3	3	NOV.	NOV.					
	3	3	DEC.						[1962]
Pattern 9	X	Y	X/Y					X	Y
Pattern 20							X	X	
Pattern 21			X					X	
Pattern 22				X				X	
Pattern 23					X			X	
Pattern 24						X		X	

volume. The computer then uses its base data to note that the last issue of this volume will be no. 7,323 + (26-1) [i.e., no. 7,348] and that the last issue for the year will be 7,323 + (52 — 1) [ie., no. 7,374] and stores this information. For the next expected issue, since the numbering system is continuous, the last interval (52) for the year has not been reached, and the last issue number for this volume has not been reached, a *1* is added to issue number and interval to prepare the next check-in card (vol. 1, no. 7,324, 1964, interval 2). This procedure continues, triggered each time by input of the received issue card, through vol. 1, no. 7,348, 1964, interval 26. On checking its stored data, the computer recognizes this as the last issue for the volume. Since the interval has not reached the stored value of the final one for the year and this is PATTERN 4 (volume numbering repeating through a 1-2 cycle each year), the volume number for the next expected issue is incremented by one. The issue number is also incremented by one, again because in PATTERN 4 issue numbering is continuous, and the interval is increased by one.

The next expected issue check-in card now reads vol. 2, no. 7,349, 1964, interval 27. Meanwhile, the stored base data are updated to indicate that the last issue of this current volume will be no. 7,348 + (26 — 1) (i.e., no. 7,374). The check-in card production continues as previously, with unit incrementing of issue and interval through vol. 2, no. 7,374, 1964, interval 52. When *this* card is fed back into the computer to indicate receipt, it is recognized by the computer to be the last issue of the volume *and* the last interval of the year. Working to instructions for PATTERN 4, the computer forecasts the next expected issue as vol. 1, no. 7,375; it also updates the year (because the final interval has been reached and the last issue of the year checked) and resets the interval to *1*. The next expected issue check-in card then reads: vol. 1, no. 7,375, 1965, interval 1. The base data, meanwhile, are updated to indicate that the last issue of *this* volume will be no. 7,400 and the last issue number for *this* year will be 7,426, and the cycle continues.

For the greatest possible brevity with a sufficiently complex pattern, this one specific example has been given. It should be evident herefrom how updating of continuously numbered volumes, repeating issue cycle numbering, and multiple-year coverage can be done. For *flop-over* volumes, the inclusion of base data for *both* the last issue of a volume *and* the last issue of the year indicate that year-updating (triggered by the interval value) can be done independent of volume number change. It should be noted that the *last issue number for the year* is *not* always synonymous with the last interval: thus in a monthly serial having, say, two volumes per year each of 6 issues, *no. 6* may be the number of the last issue of the year, but will also be a number of an issue of the previous volume *six months* earlier. The interval device is then the only year-change determinant.

• Discussion

All automated (computer- or punched-card-oriented) serials records controls are based on the predictive nature of serials issuance. Such a bald statement may make anyone who has worked closely with serials smile: for many years it has been axiomatic that the only predictable thing about serials is their unpredictability. Happily this *axiom* has been based on the true vagaries of a small number of titles, and the seeming vagaries of many. The *seeming vagaries* relate, in practice, to the already mentioned *under-age* and *over-age*, to changes of title which do not interfere with predictability of issue, and to changes in frequency, which again do not affect the basic pattern. Where there is a true change from one publication pattern to another, it may be assumed that such a change will persist for a reasonable time, and the title is merely coded under its new pattern.

The publication of two issues with the same numbering, erroneously repeated volume numbering, etc., can be dealt with by the simple literary device of labelling such issues *bis*. A simple programming device permits a distinction to be made between *bis* and *non-bis* numbers, without upsetting sequential predictability. Such anomalies disappear from the record as further correctly-numbered issues are received and the holdings record updated. If necessary, a *History note* such as "Volume 20 repeated in numbering" can be added. Supplements, when issued, are treated in one of two ways: either they form a separate series, with its own publication pattern, and can be treated as a separate title, or they form no predictable pattern and are issued sporadically as supplements to specific issues or volumes of the main title. In the latter case they can be handled as special issues, labelled *sup*, within the major pattern, being programmed in much the same manner as *bis* issues. It will usually be necessary in such sporadic supplement cases to add the information note: "Some issues have supplements." So-called *special issues* published outside the regular publication pattern numbering system are considered identical with supplements.

The problem of subdivisions beyond the *issue* level seems at first an insurmountable obstacle to the use of the publication patterns listed. Indeed, were our definition of a serial to include publications of the German *Handbuch* type, the patterns would need extensive expansion; any system which includes *works-in-parts* will find it necessary to develop patterns including the *subissue* numbering phenomenon. In our own experience, this has not been necessary.

Three distinct types of *subdivision of issue* have been found. The most frequent has been the annual-report type of publication, issued in regularly determinable 2 or more parts. Such *subdivisions* can be read as a recurrent numbering system within a serial issued in a fixed number of issues per year (i.e., as PATTERN 10), and present no problem. The second type occurs when a two-part issue

is occasionally published in a regularly volume-and-issue-numbered serial. These can be treated exactly as *bis* or *sup* issues, without creating any difficulty in pattern predictability or in the merging of the holdings record upon updating. Alternatively, such dual issues can be considered as a single issue only for purposes of the record: this procedure is justifiable in the light of both parts in fact forming only one bibliographic issue, both parts usually being issued at the same time and indeed being received in the same mailing wrapper.

The third type of issue subdivision presents a thornier problem. This case comprises the infrequent issuance of sub-issue *parts* within a continuously numbered, irregularly-issued title (PATTERNS 7 and 11). Such infrequent parts-of-issues often are published with long time periods between consecutive parts, and many times with single-issue publications intervening between their appearances. It has not been found feasible to develop a separate pattern for such publications, and they are considered as an unhappy case of PATTERN 26.

The system under development at UCLA uses publication patterns to project the next expected issue of a serial, as well as to prepare forms for bindery and to update boh bound and unbound holdings inventories. Although other systems may ignore the problem of bindery, the problem of predictability of future issues is still central, especially when more than one future issue per title is forecast for receipt as is current practice in those systems known to be operational. However named or defined, publication patterns of serials are an essential basis for computer control of serials records.

Bibliography

For those interested in basic readings on computerized serials systems, the following publications are suggested. This list is not intended to be exhaustive, but merely to give an idea of the work currently being carried on in this area of library mechanization.

UNIVERSITY OF CALIFORNIA, San Diego, Library. "Final Report: Serials Computer Project," 1964 (Unpaged mimeo.).

HAMMER, DONALD P. "Automated Operations in a University Library—A Summary," *Coll. & Res. Libr.* 26: 1965, pp. 19–29 & 44.

MOORE, EVELYN A., and BRODMAN, ESTELLE [Communications to the editor]. *Bull. Med. Libr. Assoc.* 53: 1965, pp. 99–101. (Concerns later developments of project discussed by Pizer et al.)

PIZER, IRWIN H., FRANZ, DONALD R., and BRODMAN, ESTELLE. "Mechanization of Library Procedures in the Medium-Sized Medical Library: I. The Serial Record," *Bull. Med. Libr. Assoc.* 51: 1963, pp. 313–338.

SRYGLEY, TED F. "Serials Record Instructions for a Computerized Serial System," *Libr. Resources & Tech. Services* 8: 1964, pp. 248–256.

BRIEF COMMUNICATIONS

A Note on Value Functions

The information problem could be greatly alleviated by rejecting from the store all items which failed to meet minimum criteria of probable use and organizing the remainder so that efficiency of retrieval iso ptimized. Hayes[1] refers to this concept as "activity organization."

This concept is actively used on a de facto basis in the selection policies of libraries toward books and periodicals and the storage of fast-moving documents in positions which are convenient to hand. However, the process has never been formalized.

Hayes conceives a probability distribution function $p(x)$, or "probable activity" which provides a measure of the number of requests for the document x. The most important parameter of such a function will certainly be the subject field of the document. If, however, the subject is narrowly defined so as to exclude it as a variable, what then are the independent variables of interest?

A Case Institute of Technology study[2] postulated a large number of properties intrinsic to a document which might be used to characterize its "value." These included several obvious properties—length, number of citations, journal in which published—and several less obvious properties, including number of mentions in independent lists and frequency of request for the document, which are not properties in the strict sense. These investigators found that scientists refused to render evaluations except within their own narrow specialties, which may indicate that scientists' evaluations are not feasible checks on such functions.

To test the hypothesis that documents possess intrinsic characteristics which characterize their value, the author encoded 231 articles on a single subject, from a single year, into a machine-searchable linear file, with the following items of information for each article:

Accession number	
Authors	Number of references
Journal	Number of mathematical relations
Date	Number of illustrative items (figures, tables)
Length	Length of abstract

Also encoded were index terms arranged in consistent hierarchical fashion. The articles were chosen from several prominent journals of chemical engineering, on the subject of organic chemical processes, for the year 1962. Index terms were chosen in accordance with the Chemical Engineering Thesaurus of A.I.Ch.E.

The numerical data (length, etc.) were chosen to include all obvious properties of the document which might affect its probability of selection, i.e., its "value." For example, a long article will be preferred to a short article in the usual search, simply because the information content is higher. However, many long articles are hodgepodges of information, with little specific information on single subjects. Accordingly, the value might not be a simple function of length.

The computer program employed was a straightforward term-matching routine in ALGOL 58, suitable for the Burroughs 220. In case of a match, the program entered a computation routine which computed several simple functions of the numerical data. The printout presented numerical data and computed functions in parallel with accession numbers of the articles found in the search. Three searches were performed in all; the collection was not large enough to conduct more than a very few searches.

Evaluation of the results indicated that no item of raw data (e.g., length) correlated with the relative value of the documents retrieved in any particular search; however, one computed quantity—the number of illustrative items normalized by the length—correlated very well with the adjudged value. This could be regarded as an efficiency factor, on the premise that illustrations convey information more efficiently than test.

In extensions of this work, we hope to use a larger file to relate value assignments to functions of the text: density of terms, hierarchical relationships, etc.

LAURENCE W. ROSS
Georgia Institute of Technology
Atlanta, Georgia

References

1. R. M. Hayes, "The Mathematics of Information Systems," Notes for a short course presented at the Georgia Institute of Technology, December 3–7, 1962. 135 pp.
2. Case Institute of Technology, *Measurement of Value of Recorded Information.* Final Report to NSF. Cleveland, Ohio (July 12, 1961), 93 pp.

The Chemical Abstracts KWIC Index

KWIC indexing is primarily of usce in keeping abreast of current publication.[1] *Chemical Abstracts* KWIC index has been constructed as a current awareness tool since the numbers of abstracts in its various sections grew so large that an attempt to read any section was found to be excessively time consuming. If the editors of *Chemical Abstracts* had intended the KWIC index to serve as a retrieval tool, they certainly would have found it expedient to cumulate the indexes from each issue and replace it only at the end of the year with the subject index or, on the other hand, allow the cumulations to replace the annual subject index. This latter alternative was apparently rejected based, one might surmise, on the possible inadequacy of a permuted title index versus a subject index.

This then allows us to conclude that the *Chemical Abstracts* KWIC index was not constructed as a retrieval tool and KWIC was not cumulated since it was not to serve the function of a retrieval tool even temporarily until the annual subject index was to be published. Therefore, why was it constructed? It would follow that it was constructed as a current awareness tool.

Chemical Abstracts itself is a current awareness tool and since the numbers of abstracts were so great and were thus impossible to cover even in a reasonable sized section, KWIC was introduced to make up for this deficiency. Now how, in fact, did one use or should one have used the abstracts? One turned to one's own section or sections and either glanced through the titles or read both the titles and the abstracts. One is not, in other words, looking for anything specific, but is going through the abstracts to be kept currently aware of the recent literature. Since in a KWIC index key words are scattered throughout the index one should be searching for a key word, but in fact to do so would defeat the current awareness function which means covering all the abstracts or titles on one's section or sections. One is not then trying to retrieve information; one is trying to keep currently aware of recent publication. The KWIC index should then be treated in the same way as going through the section; that is, by reading the KWIC from A to Z, or if one insists on using the KWIC as a retrieval tool, then its current awareness function is lost and reading through one's own section of the *Abstracts* in addition to searching for key words is imperative. During a year's publication, this will mean one searching through 23 KWIC indexes. Obviously this is not the intention of the KWIC indexing service; it was meant as a current awareness tool. Since this is so, how long does it take to read the KWIC index A to Z for the sake of being kept currently aware and to save time over covering the abstract section? It takes on the average of two and a

half hours per issue, and to read an average abstract section also takes roughly two and a half hours. And since KWIC is becoming larger in the neighborhood of five times faster than the abstracts, we may soon need an indexing service for KWIC. And of course if clue words are to be added to KWIC the size of the index will increase even faster. A simpler current awareness arrangement might be to abandon KWIC and to have listing by title for each of the 70 sections of *Chemical Abstracts*, or better yet to merely read the abstract titles as found in one's section of interest and to read the abstract as interest arises. It should be added that cumulating KWIC would hopelessly lose the concept of current awareness.

This brings us naturally to the question of the validity of an abstracting service when it gets to be the size of *Chemical Abstracts*. Clapp[2] neatly posed the broad question of the validity of abstracts; I believe his observation applies particularly well to the large abstracting services. Certainly *Chemical Abstracts* is the finest representative of a large abstracting service. Since the bi-weekly function of the abstracting service is to serve as a means of keeping in touch with the latest publications in one's area of interest, size itself would pose a limitation since one has only so much time to devote to abstracts, and it is well known that *Chemical Abstracts* is getting larger. The question is, when does its size reach such proportions that it can no longer be utilized as it was designed, as a means of keeping abreast of current chemical publication in one's field? I think the answer to that question is: now—or else the editors of *Chemical Abstracts* would not have considered the KWIC index as a necessary approach to the solution of the continued growth of *Chemical Abstracts*. And since *Chemical Abstracts* continues to grow it appears to me that from here on *Chemical Abstracts*, as divided now into 70 sections, will become increasingly difficult to use as a method for keeping abreast of recent chemical publication. A solution proposed by Charles Bernier that the number of sections be increased on first sight is definitely a likely approach but it is clearly recognized that the narrower one makes the subject division the lower will be relevant coverage of one's area of interest.

Chemical Abstracts to recent times has been a magnificent and successful attempt to keep up with new developments in chemical research but the literature is growing so rapidly that the literature search for new developments is undoubtedly increasingly difficult and if it could be covered, the research workers would not be able to devote the required amount of time to the increasing numbers of abstracts to keep abreast in this manner. Therefore it would appear that from here on *Chemical Abstracts* as a means of keeping currently aware of recent developments is self defeating. Perhaps an indexing service devoted to new developments in chemical research might be a partial approach to the problem. At present, I am not convinced that *Chemical Titles* would be particularly valuable as a means of keeping abreast of new chemical developments because of its vast size; it is better designed as a method for retrospective literature searching. I think, if titles are accepted as a valid source of information, that a bi-weekly listing of current chemical titles within the 70 chemical sections might be a useful expedient and this approach could be coupled with a cumulating subject or KWIC approach for further information leading to the article via the current chemical titles listed. That is, the subject indexer would be working from the article in preparing the index and making reference to the appropriate titles in the bi-weekly current chemical titles list proposed in place of abstracts. All the various *Chemical Abstract* indexes would continue to be valuable. Since neither the KWIC nor the subject index would be serving a current awareness service they should be cumulated.

The procedure for use would be as follows: a chemist would turn to one of the 70 sections of interest, read through the titles in order to keep currently abreast of recent publications of interest, and in addition he would be able, by means of the cumulating subject index, to search for subjects of current interest as well, and secondly, be able to make retrospective searches throughout the year. Since the subject indexing is taken from the substance of an article, synchronous use of the title and subject index could give an idea of the relevance and usefulness of the article to one's own research.

JAMES VAN LUIK
University of Wisconsin Library School
Madison, Wisconsin

References

1. Charles P. Bourne, *Methods of Information Handling*, New York, Wiley, 1963, p. 17.
2. Verner W. Clapp, *The Future of the Research Library*, Urbana, Illinois University, 1964, p. 85.

An Experiment Comparing Key Words Found in Indexes and Abstracts Prepared by Humans with Those in Titles*

At the Graduate Library School of the University of Pittsburgh† an experiment was conducted which aimed to provide a basis for measuring the relative effectiveness of Key-Word-in-Context (KWIC) title indexing and human indexing as practiced by members of a class in documentation offered by the school.

At the outset, each member of the class was requested to select five pieces of factual source material in a field in which he or she possessed a working knowledge. The material chosen was to be periodical articles preferably on a specific, clearly defined and limited subject in order to avoid the review article with a short and perhaps irrelevant title to which the application of KWIC indexing might be relatively ineffective.

The following indexes and abstracts of the material were to be prepared:

1. A shallow index (SI) of three entries or less.
2. A deep index (DI) of seven to ten entries.
3. A shallow abstract (SA) of twenty words or less.
4. A deep abstract (DA) of fifty words or less.

No instructions were provided regarding procedures for preparing these and no information was given about the next phase of the experiment.

After time was allowed for preparation of the above, the next phase of the experiment was announced. This was to prepare a KWIC title index with key words chosen, using consistent rules to simulate the program of a computer performing the same task, and to present in tabular form the number of correspondences between title key words and those appearing in the indexes and abstracts which had been prepared previously.

The form of the tabulations was as follows: The title key words were made column headings and the rows were identified by the initials SI, DI, SA, and DA to denote *shallow index*, etc. When a title key word appeared in the SI, DI, SA, or DA, an X was placed at the intersection of the appropriate row and column. In establishing correspondences for index entries, only the initial words of the entries were used. Also, nouns in singular and plural forms and verbs in different tenses were taken to be synonymous in establishing all correspondences.

Having tabulated correspondences as outlined above, the students were then to apply the rules used for the title key word selection to the indexes and abstracts they had prepared and to identify all the key words in these regardless of whether they appeared in the title. Having done this, they were then to count up the actual number of key words in the SI, DI, etc. which *did not* correspond with those in the title as tabulated and enter this number in another column to the right of each tabulation in the proper place. This column of figures was to be considered the *Residue* and the column so headed; again, multiple appearances of key words were to be counted as one.

The object of the experiment was to determine how many *key* words a human indexer and abstractor will produce in excess of those which would provide unique entry into the material through a machine-produced permuted title index.

In working up the results, the means and medians of the residues of the SI's, DI's, etc. were calculated. The means and medians of the total number of key words in the indexes and abstracts were also calculated. In all, 101 papers were used in the calculation. The results of the residue tabulation are given in Table 1:

* The experiment was performed in a class conducted by Allen Kent in the spring of 1963.
† Now the Graduate School of Library and Information Sciences.

TABLE 1

(Sample size N = 101)

	Sum	Mean	Median	Residue as % of Total
Shallow Index				
Total significant words	320	3.2	3	59.5
Residue	191	1.9	2	
Deep Index				
Total significant words	787	7.9	8	78.5
Residue	621	6.2	6	
Shallow Abstract				
Total significant words	1,688	16.9	17	86.5
Residue	1,463	14.6	13	
Deep Abstract				
Total significant words	4,105	41.1	42	92.6
Residue	3,814	38.1	37	

In arriving at the "total number of significant words" in Table 1, the residues and correspondences for each row were simply added. This should give the number of key words in the indexes and abstracts if the work of tabulation has been properly done. It will be noted that the residues increase steadily from SI to DA as might be expected.

When these results were presented to the class, it was suggested by a student that the papers of the group which could be considered purely *technical* in content be broken out and tabulated in the same manner to determine the magnitude of the difference, if any, between the residue percentages for articles in technical journals and those of the group as a whole. It was supposed that the percentages would be substantially lower for the technical articles. This proved to be the case.

Of the whole group of papers, only 16 could be considered to be purely technical in content. These were broken out and the results given in Table 2:

TABLE 2

(Sample size N = 16)

	Sum	Mean	Median	Residue as % of Total
Shallow Index				
Total significant words	46	2.9	3	34
Residue	17	1.0	1	
Deep Index				
Total significant words	146	9.0	9	72
Residue	103	6.5	7	
Shallow Abstract				
Total significant words	222	13.9	14	70
Residue	155	9.7	10	
Deep Abstract				
Total significant words	551	34.4	35	89
Residue	494	30.9	29	

Early in the experiment the need for a normalized sample to test the consistency of application of key word selection rules was felt. Because of the impossibility of normalizing all of the papers in the time available for the work, it was decided to take a sample of 10% of the papers, normalize and tabulate them, and compare the results with those obtained from the 101 papers prepared by the student body. In doing this, every wor which was considered non-key was listed. The list was prepared by going through the sample and selecting those words which appeared to bear little or no semantic burden and then eliminating these words systematically throughout the sample.

Having done this the residues were then counted and the results averaged. A tabulation of these appears in Table 3. To facilitate comparison, the original figures from Table 1 have been put into Table 3 along with those derived from the sample.

TABLE 3

	Sum	Mean	Median	Residue as % of Total
Shallow Index				
Total significant words—Norm.	31	3.1	3	67
Total significant words—Orig.	320	3.2	3	59.5
Residue—Norm.	20	2.0	2	
Residue—Orig.	191	1.9	2	
Deep Index				
Total significant words—Norm.	72	7.2	7	83.3
Total significant words—Orig.	787	7.9	8	78.5
Residue—Norm.	60	6.0	6	
Residue—Orig.	621	6.2	6	
Shallow Abstract				
Total significant words—Norm.	142	14.2	12	81.7
Total significant words—Orig.	1,688	16.9	17	86.5
Residue—Norm.	116	11.6	10	
Residue—Orig.	1,463	14.6	13	
Deep Abstract				
Total significant words—Norm.	295	29.5	27	90.5
Total significant words—Orig.	4,105	41.1	42	92.6
Residue—Norm.	267	26.7	24	
Residue—Orig.	3,814	38.1	37	

Looking at the percentages alone would lead one to conclude that the members of the class whose papers were used did their work in a reasonably consistent manner and that the results are meaningful as they stand. What is perhaps surprising is that the class as a whole was much more restrained in assigning non-key status to words than was the author of this paper, even though he felt at the time he was doing it that he was throwing out only those words which most clearly qualified for non-key status.

This effect becomes more pronounced as one moves down the Table from Shallow Index to Deep Abstract. This is what one would expect if he had expected such a phenomenon to occur in the first place.

In conclusion, one can say that the data presented in this communication tend to show that permuted title indexing becomes relatively less effectual as a tool for retrieval of information as need for depth of analysis increases (i.e., as one moves from shallow index to deep abstract). This effect is more pronounced in the case of non-technical subject matter of the type that would appear in publications other than professional journals. Whether the use of permuted title indexing is significantly more effective in the case of professional journals has here neither been studied nor established.

JOHN P. KNABLE, II
807 Wallace Street
Birmingham, Michigan

NASA Search System Analysis Sheet

Since May 1962, the Scientific and Technical Information Facility operated for NASA by Documentation Incorporated has carried out, edited, and delivered to selected NASA contractors nearly 700 literature searches. These searches are not necessarily restricted to the collections of the Facility, which now number well in excess of 100,000 documents. Since searching of the Facility's store is done by computer, however, a request which limits itself to this universe can be delivered many times faster than one requiring more conventional techniques, such as manual searching of abstract journal indexes. An additional inducement to the requester to accept these limitations is the fact that the open-literature materials announced by *International Aerospace Abstracts* have been included on the Facility's tapes since January 1963. A machine search by the Facility, therefore, can be said at the very least to provide complete coverage of items in both *Scientific and Technical Aerospace Reports (STAR)* and *International Aerospace Abstracts (IAA)*, the two principal abstract journals in the aerospace field. At any rate, it is machine searching, whether it be the whole or a segment of any given search, that

Bibliography #: 453 Title: Use of Fluorine for Rocket Propulsion

A. Terms/Hits

a. Total Search Terms 37

b. Maximum Hits Possible 906

c. Anticipated Hits 500

B. Most Heavily Posted Terms:

	Terms	Postings
1.	PROPELLANT	2830
2.	FUEL	1765
3.	OXIDIZER	384
4.	FLUORINE	300
5.	FLUORIDE	259

C. Type of Logical Equation Specified:

_____ a. Loose. High output. Irrelevant material expected.
_____ b. Moderately loose. Some irrelevant material.
__X__ c. Moderately tight. Very little irrelevant material.
_____ d. Tight. No irrelevant material expected.
_____ e. Analog. Analog measure: _______

D. Initial Search Results:

a. Hits (Total Output = T) 441
b. Accepted Hits After Editing (Accepted Accessions = A) 379
c. Acceptance Ratio, A/T x 100 = 86.5 %

E. Auxiliary Search Results:

a. Hits (T') 10
b. Accepted Hits (A') 10

F. Reject Analysis:

a. Rejects on Initial Search 63
b. Rejects Attributed to Type of Equation, i.e., out-of-scope or marginal upon examination 62
c. Rejects Attributed to "Noise", "False Drops", etc. ______
d. Other Rejects, e.g., Indexing Errors 1
e. Total Rejects Considered
Excessively high ___ High ___ Average ___ Low X

G. Miss Analysis:

a. Misses detected, overall 2
b. How were misses detected? STAR CUMULATIVE INDEXES — PUBLISHED SUBJECT INDEXING

H. Analyst's Comments and Recommendations (i.e., search strategy, reject analysis, new terms, delete and transfer suggestions, indexing errors, etc.):

1. THE 62 REJECTS ALL DEALT WITH FLUORINE COMPOUNDS AND FLUOROCARBONS, BUT WITHOUT SUFFICIENT RELATIONSHIP TO THEIR USE IN PROPELLANTS.

2. AUXILIARY SEARCH: CHLORINE TRIFLUORIDE AND OXYGEN DIFLUORIDE (OXYFLUORIDE) WERE SEARCHED INITIALLY UNDER THEIR "PRE-COORDINATED" NAMES. THE AUXILIARY SEARCH UTILIZED (1) CHLORINE INTERSECT FLUORINE UNION FLUORIDE, AND (2) OXYGEN INTERSECT FLUORINE UNION FLUORIDE. INDEXERS SHOULD BE ENCOURAGED TO USE "PRE-COORDINATED" PROPELLANT NAMES.

3. TERM "OXYGEN DIFLUORIDE" SHOULD BE USED IN LIEU OF "OXYFLUORIDE".

4. N62-10209: DELETE POSTING UNDER "FUEL"

5. N63-10406, N64-11223: POST UNDER "PROPELLANT".

FIG. 1. NASA Search System Analysis Sheet.

has special interest for the Facility as a prime measure of the efficacy of its indexing practices and the effectiveness of its retrieval techniques. And it is the machine search portion of any bibliography which is subjected to special analysis to achieve these measures.

During the first year and a half of operation, it was customary to deliver to the Manager of the Reference Department for final approval and release the following: 1. one copy of the original requesting letter or memorandum; 2. one copy of the analyst's work sheet, showing terms used, the number of postings against each term, and the logical equation specified; 3. the finished search itself, a machine printout of bibliographic citations. Upon approval, the search itself was transmitted with a covering letter and the other documentation was returned to the files.

With experience, however, the analysts became increasingly sophisticated in developing clever search strategies and in analyzing their results. Analysts deliberately cast their net over greater or lesser areas, depending on the stated or implied desires of the requesters to receive comprehensive searches or to receive only the most important references, dealing exclusively with the subject in question. As a result of reject analysis, feedback to the indexers became a regular feature of daily operations. This took the form of recommendations for new terms to be added to the vocabulary, recommendations that existing accessions be re-indexed to certain terms that were not used in the first indexing, recommendations for the deletion of existing terms and/or the transfer of existing postings, suggestions for the proper use of existing terms, and notification of faulty postings that existed due either to keypunch error, human error, machine error, or plain bad indexing, and which should be removed.

For these reasons, and for other administrative reasons, it became necessary to devise a paperwork instrument by which to formally communicate such recommendations to the Processing Department of the Facility. It was also necessary that management or supervision approving the final search product be provided with additional information on which to base their approval both of the product and of the work of the analyst. It was also thought that such an instrument could profitably be utilized to gather a few statistics that otherwise were widely dispersed in the files.

The *NASA Search System Analysis Sheet*, of which a fictitious completed example is presented (Fig. 1), was somewhat hurriedly designed and put into use to serve the above functions. Other organizations, engaged in or about to engage in machine searching, may find this form instructive or suggestive. It is serving its purpose at the Facility and the general consensus is that an examination of these forms provides a real "feel" for how the machine searching activity is going as a whole, and how individual analysts are progressing in particular. The form is not only passed between Departments but between analyst and analyst, analyst and indexer, and is much used as an instructional device for new employees.

This essentially prototype form can be expected to undergo considerable revision in the near future, as the Facility converts its files from an inverted structure to a linear structure, adding in the process considerably more searching and sorting versatility to the bibliography programs.

W. T. Brandhorst
P. F. Eckert
Documentation Incorporated
Bethesda, Maryland

Letter to the Editor

Dear Sir:

For several years, courses in Technical Information have been taught in Sweden, but this Fall, something new was tried. A course in *Informatology* was started at the Royal University of Technology (KTH) in Stockholm. The title of this 100-hour course has its origins in Sweden. As far as Dr. Björn V. Tell, Librarian at KTH, knows, informatology was first mentioned in a lecture given on March 8, 1962. The intention was to avoid the term *documentation,* which connotes attention to documents. Now that the emphasis in our field is on information, it was considered appropriate to coin a term with that connotation.

The term *informatology* denotes that work and research within the field of information processing which combines intuitive and algorithmic procedures where computers are used as a supplement to the human intellect. Treatment of subsets of intellectual activities which do not depend on intuition are in the domain of the *informatologist.* He would study an activity, such as classificating, for example, which generally does not suppose the use of computers or logical methods, in order to study its intuitive character and ascertain whether such activity (a series of intuitive sets) can be formalized and defined as subsets of the larger system which would employ computers. The goal of the informatologist is *data* processing and development of know-how about when and when not to use the computer. Someone trained in informatology could serve as a director of information in an institution or industrial firm. He would be equipped by his training to make decisions about new methods and products to be used for information processing in the same way that a director of research in the firm would decide on new research projects.

The new course, really an introduction to informatology, is called "Transmission of Scientific Information." It was arranged by the Swedish National Committee for Documentation, under the general direction of Dr. Tell. The outline of the course describes the background and informatological aspects covered. Of the sixty participants from industries, universities, and libraries, more than half are trained engineers and scientists.

TRANSMISSION OF SCIENTIFIC INFORMATION

Outline of Course

1. Introduction	
Basic concepts	2 hours
The logical basis for information transmission	10 hours
2. Patterns of behavior in information retrieval	
The readers' habits	2 hours
The active research worker	4 hours
Information retrieval in an enterprise	2 hours
3. Development and growth in scientific documentation	2 hours
4. Sources of information	
Primary, secondary and tertiary publications	4 hours
Institutes and organizations	2 hours
Organization and costs for information organs within an enterprise	2 hours
Patents	3 hours
5. Media	
Rules and standard for scientific authorship	4 hours
Reprography	4 hours
Translations	2 hours
6. Analysis of information	
Classification	4 hours
Coordinated indexing	6 hours
Summaries, abstracting	2 hours
7. Experimental methods for information retrieval	
Edge-notched cards, punched cards, etc.	4 hours
Computer aspects	4 hours
8. Informatological aspects	
Codes and error correction, etc.	2 hours
Neural networks, self organization and automata	2 hours
Uncertainty theories, "redundant information," etc.	2 hours
Learning and problem solving, etc.	2 hours
Theory of automata	4 hours
Linguistic problems in information retrieval	4 hours
9. Description of existing systems	10 hours
10. Exercises on a computer	10 hours
11. Seminars	4 hours

P. Atherton
American Institute of Physics

BOOK REVIEWS

4/65–1R **Language and Information: Selected Essays on Their Theory and Application.** 1964. Yehoshua Bar-Hillel. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, Palo Alto, London, and The Jerusalem Academic Press Ltd., Jerusalem, Israel. $12.50.

In the field of documentation, Bar-Hillel is known as the outstanding critic of computer applications in information processing. In the scholarly world at large, he is also known as a logician and philosopher of note. For those who are not familiar with his career, the introduction to the volume under review gives a survey of the development of his thinking which serves as the connecting link between the papers and represents a self-critical evaluation.

The book is divided into five parts: Part I. Theoretical Aspects of Language; Part II. Algebraic Linguistics; Part III. Machine Translation; Part IV. Semantic Information; Part V. Mechanization of Information Retrieval.

From this reviewer's viewpoint, however, Bar-Hillel's work can be subsumed under three major headings: 1. contributions to formal information theory; 2. algebraic linguistics; 3. a critique of machine translation and information retrieval.

In the field of information theory, Bar-Hillel's main contribution is generall recognized to be a joint paper with Carnap, "An Outline of a Theory of Semantic Information" (pp. 221–274).

Bar-Hillel's work in algebraic linguistics begins with an attempt at some original theorizing. This is represented particularly by his papers, "On Syntactical Categories" (pp. 19–37) and "A Quasi-Arithmetical Notation for Syntactic Description" (pp. 61–74). It is an attempt at introducing some logical concepts into the treatment of natural language syntax. His more recent work in the field abandons the quest for originality and is instead a very sincere attempt at understanding and further developing the syntactic theories of Noam Chomsky.

Workers in information theory and algebraic linguistics think highly of Bar-Hillel's contributions. This reviewer has no quarrel with their judgment but would like to point out that, for the field of documentation, the basic question is not that of the value of Bar-Hillel's work. Rather, it is the much deeper question of the relevance of the formal trend which he represents to the problems of the processing of information coached in a natural language. This question in turn can be answered only in terms of the very general problem of the relation of empirical method to formal theorizing in an emergent field of knowledge. In the present context, all that can be done is to raise the question, since to do it justice would go far beyond the bounds of a simple book review.

Bar-Hillel's views on machine translation and information retrieval are well-known to be negative. In regard to machine translation, he does not believe that fully automatic, high-quality machine translation is a reasonable aim. In regard to information retrieval, he does not believe that present attempts at theorizing and computer processing are very useful.

If one takes feasibility and utility in the very strict sense in which Bar-Hillel uses these criteria, one can, of course, agree with him. Translation of the quality produced by a good literary translator will be difficult to achieve automatically, if at all. Much of the work in the theory and computer applications of information processing is quite naive and certainly not as useful as it could be.

In this reviewer's opinion, however, Bar-Hillel asks the wrong question. The problem is not whether it is logically possible to conceive of a perfect scheme. To that question, the logical answer is negative, and Bar-Hillel marshalls a number of good reasons why this is so.

With respect to feasibility, this reviewer's preference would be to ask a more modest question: What is required in order to make progress in the two fields, and how can this requirement best be met? As far as usefulness is concerned, this is largely related to the degree of need. Bar-Hillel denies, in "Is Information Retrieval Approaching a Crisis?" (pp. 365–372), that there is an information explosion requiring the mechanization of information processing. He seems to forget that even if—as he claims—scientists may be happy enough with their specialization and word-of-mouth transmission of information, there is still the need to meet the enormous demands of the research administrators and other specialized interests in government and elsewhere.

In spite of the disagreement with the essence of Bar-Hillel's argument, this reviewer finds that in the final analysis, his negativism has had a positive effect on the field of documentation. It has forced an increased awareness of the difficulties and it has influenced at least some workers to take their theories less seriously.

A book of selected writings is always a great convenience. In the present case, the figure of Bar-Hillel as a scholar emerges with all his virtues and flaws not only through the papers themselves but also through the most informative introduction.

Paul L. Garvin
The Bunker-Ramo Corporation
Canoga Park, California

4/65–2R **Subject Authority List.** January 1964. American Petroleum Institute Information Retrieval Project, Central Abstracting Service, 1271 Avenue of the Americas, New York, New York 10020. Looseleaf. 245 pp. plus 74 pp. in Appendix and Hierarchy.

For several years the American Petroleum Institute (API) Central Abstracting Service, under the direction of Everett H. Brenner, has published *Abstracts of Refining Literature* and *Abstracts of Refining Patents*. These two publications are now supplemented by an Information Retrieval System which includes: 1. Alphabetical Subject Indexes to the above-mentioned publications in book form on a monthly basis; 2. Dual Dictionary Coordinate Indexes produced by computer; 3. Abstract Cards; 4. Computer tapes and programs for mechanized searching; and 5. The backbone of this system, the *Subject Authority List* (SAL).

This basic tool is intended to be used by both the indexer and searcher using manual or machine methods to cover the 25,000 papers and patents in the open literature on oil refining produced

annually. Three and a half years were spent in developing SAL and the new system. It was a joint effort of the API staff and representatives from the more than 20 oil companies served by API's Central Abstracting Service.

The entire system is not under review here, but SAL as a publication is best assessed in context. It was not produced as an idle exercise by someone who wanted to create the *best* indexing language ever devised. It was designed as a working tool to meet the special needs of several organizations and its characteristics were partly dictated by the system of which it is a component. SAL's format, looseleaf, is indicative of its dynamic character. A second edition is planned for early 1965.

A rundown of the contents of SAL exhibits its hybrid characteristics, combining the elements of a library subject heading list, a *dictionary* for mechanized coordinate searching with role indicators, and a classification scheme or a thesaurus, with group designations:

I. *Alphabetical Thesaurus* (each descriptor complete with scope notes, "use" cross references, listings of narrower and broader terms, "see also" references).

II. *Chemical "Aspect"* words to characterize structural features of chemical compounds.

III. *Hierarchy* (with *Index to Hierarchy* for quick scanning) for all descriptors, cross references, broader term/narrower term relationships and "see also" references are presented here under broad headings:

(a) Administration
(b) Basic Science
(c) Properties
(d) Equipment
(e) Material
(f) Processes
(g) Miscellaneous

IV. *Information Aspects* for characterizing document (e.g., catalog, book, review, specifications, etc.).

V. *Descriptors for Patents Only.*

VI. *Role Indicators* (for reactions only: agent, product, catalyst, carrier, prior treatment).

It is reported that SAL is strictly followed in the API System. This facilitates the multiple uses made of the several products produced by the Central Abstracting Service. Apparently, every effort was made to take the best of the parent systems from which this tool developed and to avoid their limitations. Multiple-word subject headings, for example, have been avoided, but the uncontrolled vocabulary of pure coordinate systems is not allowed. The *Hierarchy* formed the basis for the broader/narrower relationships, but it is not the rigid hierarchical authority list which some classification schemes have become. The *Hierarchy* is used when a new cross reference or descriptor is added so that the relationship of the new to all other descriptors can be shown.

The *Alphabetic Thesaurus* does not indicate the placement of descriptors in the *Hierarchy*. This is one feature of traditional subject heading lists (L.C. and Sears) which it might have been well to include. To use the *Hierarchy* for specific reference, you need to know, e.g., that "Absorption Processes" and its broader term "Sorption Processes" is under *"Processes/Physical Processes/Separation"* in the *Hierarchy*. The *Index to the Hierarchy* (really a contents page in very small print) is not very helpful. Some form of ordinal notation or location symbol would be helpful. Only the thesaurus is a computer printout. If some notation were added automatically to both the thesaurus and the hierarchy, they could both be linked and updated by computer.

There is no guide for use of SAL published in SAL. I understand this may be corrected in the next edition.

Continuous revision is planned, but it is hoped they will try some easier way of correcting old editions than marking additions and changes by hand.

Paul Howerton, at the 1963 ADI annual meeting, said that this is about "the-sauri-est" time in the history of documentation. No doubt, he is right, but this publication of API is one of the best published in many a year, from the standpoint of general features. Some specialist in petroleum literature will have to assess the specific descriptors and their relationships, as this reviewer did not feel qualified to do so.

PAULINE ATHERTON
American Institute of Physics

4/65–3R **An Evaluation of Links and Roles Used in Information Retrieval.** Thesis Presented to the Faculty of the School of Engineering of the Air Force Institute of Technology, Air University. 1963. Lieutenant Colonel Jefferson D. Sinnett. Air Force Institute of Technology, School of Engineering, Wright-Patterson AFB, Ohio. AD 432198.

The author carried out an experimental investigation of the efficacy of links and roles in a retrieval system being developed by the University of Dayton for the Air Force Materials Laboratory, Wright-Patterson AFB, Ohio. The system is an inverted coordinate index on magnetic tape, searched by means of an NCR 304 computer. The test was undertaken at a time when there were approximately 6,000 documents in the collection, dealing with adhesives, ceramics, cermets, coatings, fuels, lubricants, electrical and electronic materials, fibrous materials, metals, oils, plastics, polymers, and various types of manufacturing procedures and methods.[1] A thesaurus of about 18,000 multidisciplinary terms is used as the authority list. Relationships between terms are expressed by means of a specially-devised set of 16 role indicators.

The investigation was effected through an analysis of the results of 22 searches. The questions used were mainly synthetic questions based on documents known to be in the collection. A few, however, were *real-life* questions.

Search programs were prepared for these test questions. This entailed a conceptual analysis of the queries, translation of the concepts into descriptors selected from the thesaurus, addition of role indicators, and specification of the logical relationships (sum, product, negation) required among descriptors. The programs were then coded (numerical codes for terms and roles), keypunched, and searched on the NCR 304. Each search was run in four different ways:

A Term coordinations without links or roles
B Term coordinations with links only
C Term coordinations with roles only
D Term coordinations with links and roles

The report contains some interesting data on times required for search preparation and actual searching for each method. The average number of *key terms* (representing key concepts) in the original queries was 3.7. In a complete search program, without roles, the average number of descriptors used was 18. When roles were used, the average number of role-term combinations per question was 62.5.

Formulation and processing times were as follows:

Average formulating time in minutes per question (intellectual task of preparing search programs)		Average processing time in minutes per question (translation of the search programs into a form suitable for keypunching)		Average for both procedures
Without Roles:	21.3	Without Roles:	5.7	27.0
With Roles:	45	With Roles:	19.3	64.3

In other words, it was found that use of roles more than doubled the time required at the query formulation and processing stages. It was estimated that the use of roles more than doubled time needed at the indexing stage also. Links had no effect at the query formulation and processing stages, although they did affect searching speeds:

Average Search Times in Minutes	
Method A	7
Method B	14.3
Method C	17.6
Method D	34.9

Sinnett notes that the use of the role indicators created definite problems in the formulation of search programs. These observations bear out the experiences of other investigators[2,3] who have reported that searchers were unable confidently to predict the role combinations likely to have been chosen in the document indexing. In the DuPont and Bureau of Ships studies, role indicators proved difficult to apply because they were not mutually exclusive. Moreover, it was found that lack of discrimination,

caused by overlapping role definitions, was largely responsible for the blocking of relevant information.

The following tabulation summarizes the retrieval performance of each of the four search methods for the 22 test questions:

	A	B	C	D
Total Retrieved	548	425	501	392
Estimated Relevant	361	343	344	324
Estimated Not Relevant	187	82	157	68

Relevance figures were determined in a rather unusual manner. Most of the documents retrieved in common by all four methods were assumed to be relevant. So 10% to 100% of this *common set* was added to the *complementary sets* (i.e., those documents retrieved but not belonging to the *common set*). Each document of these augmented complementary sets was assessed for relevance. All the unexamined items of the common set were assumed to be relevant. Therefore, their number was added to that of those actually examined and found to be relevant. The author himself, "with technical assistance," assessed relevance.

Method A was assumed to be a perfect *all-but-not-only* system. That is, it was assumed to retrieve all, but not only all, documents relevant to a request. Therefore, the balance of the collection was not examined by any method to find whether any relevant documents had been overlooked by all four methods.

For evaluating retrieval effectiveness, Sinnett uses a modification of a formula proposed by Borko:[4]

the retrieval effectiveness score, $R = K\ (r - i)$,
where $r = s/t$, $i = M/N$,

$K = 100$

s = number of relevant documents retrieved (R in Cranfield notation)

t = total of known relevant documents is in the collection (C in Cranfield notation)

M = number of irrelevant documents retrieved ($L - R$ in Cranfield notation)

N = total number of documents retrieved in searching (L in Cranfield notation)

The ratio r is the Cranfield *recall ratio* (R/C), although Sinnett misleadingly calls it the *relevancy score*. The ratio i is a *noise factor* which is the complement of the Cranfield *relevance* (precision) *ratio*. This ratio has been called a "laxity ratio" by Fairthorne. Sinnett's single measure of effectiveness is thus the difference between the recall and noise ratios.

The optimum figure would be 100 (i.e. 100% recall, 100% relevance)

$$\text{(i.e. } R = 100\left(\frac{s}{t} - \frac{M}{u}\right) \text{ or}$$

$$R = 100\ (1 - 0) = 100$$

The poorest effectiveness figure would be — 100 (representing zero recall and 100% irrelevance):

$$R = 100\ (0 - 1) = -100$$

Turning the search results into Cranfield-type ratios and assuming that Method A omits no relevant items that may be in the collection, we have the following results:

	Recall Ratios	*Relevance* (Precision) Ratios
Method A	$\frac{361}{361} = 100\%$	$\frac{361}{548} = 65.9\%$
Method B	$\frac{343}{361} = 95\%$	$\frac{343}{425} = 80.7\%$
Method C	$\frac{344}{361} = 95.3\%$	$\frac{344}{501} = 68.7\%$
Method D	$\frac{324}{361} = 89.8\%$	$\frac{324}{392} = 82.7\%$

In other words, if we accept the searching method without links or roles as a perfect all-but-not-only system, the results of Sinnett's investigation were: the use of links reduced recall by only 5%, while improving relevance by 14.8%; addition of roles to the lings further reduced recall by another 5.2%, while further improving relevance by only 2%; the use of roles alone reduced recall by 4.7%, while improving relevance by only 2.8%.

Sinnett worked out effectiveness figures for the four methods of search for each of the 22 questions. He then summed the effectiveness figures for each method and divided by 22 to give the *average effectiveness scores:*

A	51.7	C	49.4
B	64.7	D	57.7

The pure effectiveness ranking would be:

- B Use of links only
- D Use of links and roles
- A Coordinations without links or roles
- C Use of roles only

However, taking economics into account, Sinnett rates the systems B, A, D, C.

As a result of his study, the author firmly recommends that the use of roles should be discontinued in the information storage and retrieval system being developed for the Air Force Materials Laboratory. He estimates that this would result in approximately a threefold increase in the rate of indexing and the rate of retrieving answers. Increased emphasis should be placed on the use of stylized statements in the application of links to indexing.

This is a valuable thesis, adding to our fund of knowledge on the performance of index language devices and on the costs of operating retrieval systems with and without various refinements. It re-emphasises some of the problems involved in the application of a system of role indicators. It is misleading, however, to attempt a comparison of systems on the basis of a single figure of merit. A document retrieval system is established to satisfy the unique requirements of a particular group of users. These requirements may demand a system capable of high recall (with inevitable loss of precision) or they may demand high precision (with consequent loss on the recall side). We must gauge the effectiveness of a system by the degree to which it satisfies these user needs. To measure system performance in terms of the balance it achieves between recall and precision is not always helpful.

At the time of writing his report, Sinnett knew of the work of Borko and Bornstein, but was seemingly unaware of other studies in the evaluation of retrieval systems. In particular, he showed no knowledge of the Cranfield investigations. The test design may have been improved, and the author's work made considerably lighter, if he had taken the trouble to familiarize himself with the growing body of literature on systems evaluation.

F. W. Lancaster
Herner and Company
Washington, D. C.

References

1. E. A. Janning, *Establishment of a Coordinate Indexing Retrieval System for the Air Force Materials Laboratory*, Air Force Materials Laboratory, Dayton, Ohio, 1963. RTD–TDR–63–4263. AD 428423.
2. W. F. Johanningsmeier and F. W. Lancaster, *Project SHARP (Ships Analysis and Retrieval Project) Information Storage and Retrieval System: Evaluation of Indexing Procedures and Retrieval Effectiveness*, Department of the Navy, Bureau of Ships, Washington, D. C., June 1964. NAVSHIPS 250–210–3. (Available from the Superintendent of Documents.)
3. J. G. Van Oot and others, "Links and Roles in Co-ordinate Indexing and Searching: an Economic Study of Their Use, and an Evaluation of Their Effect on Relevance and Recall," Paper presented before the Division of Chemical Literature, American Chemical Society, Chicago, Illinois, September 3, 1964.
4. H. Borko, "A Research Plan for Evaluating the Effectiveness of Various Indexing Systems," System Development Corporation, Santa Monica, July 1961. (Field Note 5649/000/01.)

4/65–4R **Lexicon of Archive Terminology (French, English, German, Spanish, Italian, Dutch).** 1964. Jean Herbert, ed. Elsevier Publishing Co., New York. 83 pp.

This lexicon of 175 terms was compiled by a Committee of the C.I.A. (no! not that one—Conseil International des Archives under UNESCO). There is no question but that the development of multilingual technical glossaries are of great value to the practitioners of the glossary subject. There is considerable question as to the semantic correspondence among the several languages if attention is not paid to the fact that different cultures which use the same mother tongue for general communication do not do so in technical communication. The lexicon under review suffers from this *mal des langues vivantes*.

The Committee member who was the English language lexicographer was from the United Kingdom. He did not represent American English usage. A case-in-point is term 99 on p. 35—the definition of the French word "rubrique" is: "The word *Rubrique* is used to designate a group of documents or articles which deal with the same subject." The French lexicographer further explains that the term is "used only in the sense of category by French archivists." The English lexicographer indicates that the concept is not present in English language usage. Perhaps so in England, but a quick check with the National Archives personnel verified that "rubic" is used in this country in archival work in exactly the same sense as the Frenchman defined the term.

The English language lexicographer does not accept the term "protocol" as an official and authentic statement of the proceedings of an official conclave convened for whatever purpose. There are other examples of the non-agreement of the use of the same term by the two principal English-speaking countries.

The definitions of all terms are given in French with the equivalents in the other languages following. If notes are needed to qualify the terms expressed in languages other than French, the notes are also in French.

This lexicon is shallow in technical archive terminology. It will probably have greater use in Europe than in the United States. American archivists or historians doing research in European archives may find some value in the work.

Paul Howerton
Director, Center for
Technology and Administration
The American University

LITERATURE NOTES

CHARLES P. BOURNE and BURTON E. LAMKIN, Editors

Contributors of abstracts from readers and suggestions of books and articles for review or inclusion in this bibliography will be welcome and are actively solicited. Volunteer abstractors and reviewers are needed, as well as people with special linguistic talents. All copies of reprints, reports, and correspondence for this section should be addressed to Mr. Burton E. Lamkin, 1649 Fairorchard Ave., San Jose, Calif. 95125. In order to increase its coverage, this section will include some abstracts copied or adapted from other publications. The American Documentation Institute is not able to supply copies of the publications abstracted or cited.

Abstractors

The following persons have prepared some of the abstracts for this issue. Their names are listed here to express the editor's thanks and appreciation, as well as to explain the initials that were used to sign the abstracts.

B.K.D.	Bernard K. Dennis
C.D.G.	C. D. Gull
C.L.B.	Curtis L. Brown
D.C.W.	David C. Weber
F.M.	Fenster Moop
J.A.W.	Judith A. Werdel
J.J.C.	J. J. Cirrito
P.A.R.	Phyllis A. Richmond
P.M.B.	Peter M. Bernays
T.G.F.	Trudy G. Fisher
V.E.Y.	Virginia E. Yagello

Contents

Starting with the January 1965 issue, separate copies of this section of *American Documentation* may be obtained as reprints on a subscription basis for $4.00 per year. Requests should be directed to ADI, 2000 P Street N.W., Washington, D. C., 20036. It is hoped that in the very near future it will be possible to provide Literature Notes as a publication separate from *American Documentation* and with increased coverage. The availability of separate reprint sections is a start toward that objective.

Index of Authors and Corporate Sources Referred to in this Issue

(The numbers refer to the terminal digits of the abstract serial numbers in this issue. The letter R refers to Reviews.)

Journal Abbreviations

Many of the journal titles are represented by their corresponding CODEN notation as defined in the two publications, *Coden for Periodical Titles,* American Society for Testing and Materials, Phila., Pa., 1963, ASTM Special Technical Publication No. 329, and *Coden for Periodical Titles, Supplement 1,* 1964, ASTM Special Technical Publication No. 329—31. The CODENS are being used in place of the full titles primarily to reduce typing and typesetting work, and to provide a more compact representation of the citation. The following table will help translate the CODENS for this issue.

ALBL	ALA Bulletin
AMDO	American Documentation
ASLP	Aslib Proceedings
BABL	Babel
BATR	Battelle Technical Review
BEHS	Behavioral Sciences
BHOR	Business Horizons
BMLA	Bull. Medical Lib. Assoc.
BRNS	Barron's
BUAT	Business Automation
CACM	ACM Communications
CENE	Chemical and Engineering News
CENG	Control Engineering
CMPG	Computer Journal
CPGR	ACM Computing Reviews
DKMN	Dokumentation
DTMN	Datamation
ERGO	Ergonomics
FEPR	Federation Proceedings
FRON	Frontier
IBMJ	IBM Journal
IEAI	IEEE Trans. on Applications & Industry
IEEI	IEEE Trans. on Ind. Electr. & Control Instr.
IEEP	Proc. IEEE
IETT	IEEE Trans. on Information Theory
IEEW	IEEE Trans. on Eng. Writing & Speech
IPAL	Inland Printer and American Lithographer

IREL	IEEE Trans. on Electronic Computers
IRTE	IEEE Trans. on Eng. Management
ISCT	Int'l. Science & Technology
ISSR	Information Storage & Retrieval
JACM	ACM Journal
JCHD	J. Chemical Documentation
JDOC	J. Documentation
LCIB	Library of Congress Information Bull.
LIBJ	Library Journal
LIBL	Library Literature
LIBQ	Library Quarterly
LIBT	Library Trends
LRTS	Library Resources & Technical Services
MTRS	Materials Research & Standards
NADO	Nachrichten für Dokumentation
NAMN	National Micro-News
NMRV	Navy Management Review
NWSC	New Scientist
OPRE	Operations Research
PRMG	Printing Magazine/National Lithogr.
REDE	Research/Development
REDO	Revue Internationale de la Documentation
REPM	Reproduction Methods for Business & Ind.
RESM	Research Management
RPRV	Reproduction Review
SCIE	Science
SHBN	Stechert-Hafner Book News
SPLB	Special Libraries
SSTM	Systems
STNW	Sci-Tech News
TAGU	Trans. of the Amer. Geophysical Union
TAPP	Tappi
UNDL	UNESCO Bull. for Libraries
WLDU	Wilson Library Bull.

Conference Proceedings and Collections of Papers

4/65–1 **Parameters of Information Science. Proceedings of the 1964 Annual Meeting of the American Documentation Institute.** October 5-8, 1964. A. W. Elias, Ed. Spartan Books, 1106 Connecticut Ave., N.W., Washington, D. C. 521 pp.; price **$15.75.**

4/65–2 **Automation and Scientific Communication. Part 3. Proceedings of the 26th Annual Meeting of the American Documentation Institute,** Chicago, Ill. October 6-11, 1963. American Documentation Institute, Washington, D. C. Price **$2.50.**

Contents: Introductory remarks by Burton W. Adkinson, Alvin M. Weinberg, and Roman C. Pucinski. Transcripts of the 12 Theme Session Discussions. Registration List. Combined Conference author index. Six formal conference papers.

4/65–3 **Proceedings of the 14th Annual Conference, Research & Engineering Council of the Graphic Arts Industry,** Rochester, N. Y. May 18-20, 1964. Available from the R & E Council, Washington, D. C. 125 pp.

Several of the papers deal with advances in graphic arts and related technologies, particularly research on optical character-reading and print-scanning systems (IBM, Sylvania, Philco, and Farrington), novel typesetting machines (ZIP-Photon 900 and Mergenthaler Linotron), and xerographic and other electrostatic printing methods. (C.L.B.)

4/65–4 **Proceedings of the Technical Sessions at the Patent Office of the Federal Republic of Germany.** Second Annual Meeting of ICIREPAT (Committee for International Cooperation in Information Retrieval Among Examining Patent Offices), Munich. September 4-6, 1962. Office of Research and Development, Patent Office, U.S. Department of Commerce, Washington, D. C.

4/65–5 **Proceedings of a Conference on Information Storage, Retrieval and Exchange in the Neurological and Communication Sciences,** Los Angeles, Calif. March 4-5, 1963. National Institute of Neurological Diseases and Blindness Council, Research Subcommittee. 140 pp.

4/65–6 **Proceedings of OAR Publications Conference.** May 20–24, 1963. Office of Aerospace Research, Washington, D. C. (AD–426 679). OTS price $5.60.

OAR publications should seek to bring about the impact of science on Air Force technology on a timely and continuing basis. Therefore, OAR must be responsible not only for the conduct and support of research, but also for the dissemination of results to meet Air Force needs. Publication in scientific journals is preferable; otherwise, in technical reports, which should be of same quality as comparable articles in the learned journals. Manuscripts awaiting journal publication should be re-examined within nine months to ascertain if they should be published as in-house reports. OAR must reserve the right to use and reproduce its sponsored papers and reports. Review publications are a legitimate basic research function and justify the use of PB 680 funds. They should be directed primarily toward the needs of the Air Force in-house exploratory research. Primary distribution should be kept to a minimum. Manuscripts should not be formally distributed; however, they should be promptly announced, and if appropriate, may be circulated as working papers to meet Air Force needs. (author)

4/65–7 **Proceedings of the First USAF Scientific and Technical Information Conference.** Dayton, Ohio. September 30-October 4, 1963. A. G. Hoshovsky, Ed. Office of Aerospace Research, USAF, Washington, D. C. 157 pp.

The broad objective of the conference was to increase the effectiveness of the scientific and technical programs of the U.S. Air Force. The more specific objective was to examine and suggest improved means for scientific and technical information flow. Efficient, broad interchange of pertinent, timely, and precise technical information is necessary for the advancement of Air Force technology.

Members of the conference suggested that a general approach to definition of the requirements could best be made by classifying the users as having a vital need and being supplied with fully detailed documents, or as having a casual need and being supplied with only abstracts of basic information documents. Stress must be placed on the quality of information input. The author and his organization must objectively judge the values in the work reported and discard non-essentials.

The conference discussed the protection of the copyright and patent interests of those contributing to a broad scientific and technical information interchange system It was the conference's conclusion that governmental agencies must assume proper responsibility for safeguarding proprietary information contributed to the interchange system. It was also indicated that legislation and administrative practice need to be reformed to avoid excessive limitation and delay occasioned by the contractor's proprietary interest. No specific plans or suggestions to overcome the present defects in these connections, however, were presented.

4/65–8 **Lectures on Documentation.** Björn V. Tell. A collection of papers presented at a UNESCO Regional Training Course in Scientific Documentation for Asia, New Delhi, 1963. Royal Institute of Technology, Stockholm, Sweden.

Contents:

Communication of Science Information.
Development of the Concept of Documentation.

Abstracting.
Communication of Knowledge—Sources of Information.
How a Documentation Centre Can Help Scientific and Engineering Communication
Introduction UDC.
Introduction to the Use of Punched Cards in Documentation.
Introduction to Nuclear Science Documentation.

4/65–9 **Information Storage and Retrieval.** August, 1963. Report ISR–4 of the Computation Laboratory, Harvard University, Cambridge, Mass. AFCRL–63–366. (AD–422 667.)

Contents:

Vector Images in Document Retrieval. Paul Switzer.
A Method of Maintaining Optimality in Binary Search Trees. Dean S. MacLaughlin.
Some Storage Techniques for Use with Disk Files. Franklin T. Baker.
Some Experiments in the Generation of Word and Document Associations. Gerard Salton.
Automatic Structure—Matching Procedures and Some Typical Retrieval Applications. Gerard Salton and E. H. Sussenguth, Jr.

4/65–10 **Information Storage and Retrieval.** January 1964. Report ISR–5 of the Computation Laboratory, Harvard University, Cambridge, Mass.

Contents:

A Flexible Automatic System for the Organization, Storage, and Retrieval of Language Data (SMART). Gerard Salton.

An automatic information retrieval system programmed for operation on the IBM 7094 is described. It is designed to process English texts and search requests, and to identify items of information similar to each other according to certain specified criteria. Optional features include the possibility of replacing the original text by hierarchical concept numbers, of computing term correlations using co-occurrences of terms within the individual sentences of a given text, of performing a syntactic analysis and using structure-matching procedures to compare the various texts, and of computing both term and document clusters using term co-occurrences within the documents of a given collection. The various features are controlled by a supervisory program called CHIEF. The system is designed to operate in an environment of man-machine interaction; moreover, by comparing the results of a retrieval operation executed with several different search procedures, the efficacy of the various procedures can be evaluated. The system may also be useful not only as an operating retrieval system, but also as a tool for deciding on an optimal sequence of request alterations and for judging the adequacy of the several statistical, syntactic and semantic features (author).

Automatic Identification of Phrases for Document Classification. Alan Lemmon.

Many types of information retrieval systems rely upon word frequency counts to characterize documents. This approach tends to ambiguity and better results may be obtained by counting phrases insted of words. This section describes a working system which uses a syntactic analyzer and a structure-matching routine to identify actual phrases (author).

Automatic Structure-Matching Procedures. Edward H. Sussenguth, Jr.

4/65–11 **Information Storage and Retrieval.** June 1964. Report ISR–7 of the Computation Laboratory, Harvard University, Cambridge, Mass.

The report is mostly a detailed description of the SMART experimental document retrieval system, which has been programmed at the laboratory over the uast two years and is now in operation on the IBM 7094. Section I, by G. Salton, concentrates on the system philosophy, the reasons behind the chosen organization, and the expected uses and advantages of the system. Section II, by M. Lesk, contains a general program description, including the organization of the supervisory system, (CHIEF) and of the subroutine calling sequences. Sections III, IV, and V by C. Harris, M. Cane and G. Shapiro, respectively, deal with the construction, use, and updating of the word dictionaries and of the concept hierarchy incorporated into the system. Sections VI, VII, and VIII by A. Lemmon, E. H. Sussenguth, and T. Evslin and T. Lewis, respectively, deal with the syntactic procedures.

The statistical procedures incorporated into the SMART system are described in Secs. IX and X by M. Lesk and T. Evslin, and M. Lesk, respectively. Section IX covers a set of programs designed to recognize certain concept sets known as "statistical phrases;" the components of such phrases are not related syntactically, but are defined by their co-occurrence characteristics within the sentences of the various documents. Additional statistical processing methods, including the generation of term-term and document-document similarity coefficients, and the production of term and document clusters are described in Sec. X.

Section XI, the last in this series, is concerned with the description of a variety of housekeeping routines; some of these are internal to the system, and are as such inaccessible to the user. The last three sections included in this report, No. XII by J. J. Rocchio, and Nos. XIII and XIV by A. R. LeSchack do not properly cover any part of the existing SMART systems, but deal with extensions to the system and with additional research of a statistical nature. (authors)

4/65–12 **Information Storage and Retrieval.** December 1964. Report ISR–8 of the Computation Laboratory, Harvard University, Cambridge, Mass.

This report, like ISR–7, is a description of work conducted in connection with the SMART system over the past several months. Two aspects of the system are stressed in particular: its potential as an operating, fully automatic, text processing and retrieval system; and its use as a testing device to evaluate the effectiveness of a large variety of automatic content analysis procedures.

Sections I and II by M. Lesk are concerned with a description of SMART as an operating retrieval system. Section II contains detailed operating instructions which should make it possible to prepare data and program decks, and to run the system at installations with the compatible computer and compatible monitor systems.

Sections III, IV, and X are devoted to the general problem of retrieval systems evaluation. Section III, by J. Rocchio, introduces a set of new functions to measure the effectiveness of retrieval performance. They are not dependent on a possibly artificial distinction between retrieved and non-retrieved information, as are the standard recall and precision (relevance) measures. Instead, they are based on the rank orders of the relevant documents, when the document collection is arranged in decreasing correlation order with the search requests.

Section IV by G. Salton contains the initial results of the evaluation process obtained with the SMART system over the past several months. The data reported cover approximately 20 search requests used with a collection of 405 document abstracts, and a set of about 15 different analysis procedures. It is found that methods using complete abstract texts are superior to those which confine the analysis to the titles only. Section X by J. Rocchio and M. Engel describes the evaluation process in more detail, and presents in tabular form the retrieval results obtained for each search request and each processing method used. In Sec. V, by G. Salton, the phrase generating process is discussed in detail, and it is shown how the phrase-matching methods make it possible effectively to assign the same content identifications to hundreds of semantically equivalent but syntactically quite different constructions. The "criterion phrases," consisting of combinations of syntactically related terms, are stored in the computer in a special predetermined format. The format specifications used for the construction of the criterion phrase dictionary are described in detail in Sec. VI by A. Lemmon.

The SMART system requires for its operations a set of five principal dictionaries: alphabetic stem and suffix dictionaries, a concept hierarchy in tree form, and statistical as well as syntactic (criterion) phrase dictionaries. Section VII by C. Harris describes some of the problems which arise in the original construction and in the updating of these dictionaries. Section VIII by T. Evslin

and IX by M. Cane are devoted to various extensions of the SMART retrieval system which are presently being actively considered. In particular, since the dictionary lookup routines operate in core, and require simultaneous storage of the complete thesaurus (alphabetic dictionary) and of one complete document at any given time, the size of the thesaurus as well as the permissible document lengths are strictly limited. Limitations are also imposed on the permissible size of the complete document file. These size limitations are presently being removed.

The operations of the SMART system can be adapted to an organization under user control, as is described in more detail in Sec. IX. In particular, a supervisory system is introduced in that section, as well as a set of input-output type-in and print-out statements, which are designed to adapt existing operations to the M.I.T. compatible time-sharing system (CTSS). Programming efforts in this direction are presently under way. (authors)

Published Reference Works

4/65–13 **Selected Bibliography on Indexing in Science and Technology: Theory, Application and Techniques.** Philip L. Dopkowski, Ed. Bull. No. 9, August 1963. The American University, Washington, D. C. 101 pp.

Although no computer was used in the preparation of this bibliography, it is of interest to workers in information storage and retrieval for two reasons: 1. it contains many well-annotated references on the subject of indexing, and 2. it contains two permuted indexes in which computer techniques have been simulated manually. The 141 annotated citations are arranged by type of document (i.e., book, book chapter, journal article, report, paper, etc.) and are cross-referenced by an author index (giving senior author only), a permuted title index, and a permuted descriptor index.

Much familiar and useful material can be found here; most of the items recognized are significant pieces with real content. One notable omission is the lack of any articles from the ICSI conference ["International Conference on Scientific Information," Washington, D. C. November 16–21, 1958. Proceedings. National Academy of Science—National Research Council, 1959. 2 v.].

Although the permuted title/descriptor index has been found objectionable to some, it appears to have found a deservedly permanent place in the repertoire of information storage and retrieval techniques. In the opinion of this reviewer, the value of this bibliography is greatly enhanced by the addition of the permuted indexes provided.

4/65–14 **Bibliography on Reproduction of Documentary Information.** January-February. 1963. Loretta J. Kiersky, Ed. SPLB, vol. 55, no. 3, March 1964, pp. 160–165.

Bibliography includes references concerning methods, processes, equipment, applications, systems, and standards.

4/65–15 **A Selected Bibliography of Documentation and Information Retrieval.** Autumn 1964. Western Reserve University, School of Library Science, Cleveland, Ohio. 10 pp.

4/65–16 **Documentation and Information Theory. A DDC Report Bibliography.** Bibliography for March 1961–September 1962. Defense Documentation Center, Alexandria, Va. (Rept. no. ARB 12 827. AD–447 069. Supplement to AD–269 800 and AD–267 000.) 38 pp.

4/65–17 **Department of Defense Supported Information Storage and Retrieval Program Bibliography.** February 1964, 1 vol. Reatha E. Barnes, Comp. Defense Documentation Center, Alexandria, Va. (AD–440 230.) A report bibliography.

4/65–18 **Legibility of Alphanumeric Characters and Other Symbols: I. A Permuted Title Index and Bibliography.** December 15. 1964. Douglas Y. Cornog, F. Clayton Rose, and Josephine L. Walkowicz. NBS Misc. Publication 262–1. 100 pp.; USGPO price 60 cents.

This permuted title index and bibliography to the literature on the legibility of alphanumeric characters and other symbols includes 325 references. Due to the confusion and overlap of terminology in the legibility literature, this report uses the term "legibility" to include "legibility," "readability," "perceptibility," "visibility" and any other closely related concepts. "Other symbols" include such items as the arrows and other coded symbols used to present information on radar displays. The psychological literature on perception has been included only when it was closely involved with the specific problems of alphanumeric characters and other meaningful symbols. Studies concerned with environmental variables, i.e. illumination and symbol-background contrast, and the legibility of dials and scales have received little attention. An author index is included. (authors)

4/65–19 **Thesaurus of FAA Descriptors.** ed. no. 1. July 1964. Federal Aviation Agency, Washington, D. C. (AD–604 408.) 265 pp.; OTS prices: HC $6.00/MF $1.25.

The thesaurus of FAA descriptors, compiled as part of a program for developing a semi-automated information retrieval system for the FAA, is presented. Descriptors listed were selected in the indexing in depth of approximately 5,600 technical reports. (author)

4/65–20 **Root—Term Glossary for the Subject Analysis of Technical Reports which Are of Interest to, or in the Collections of the Bureau of Aeronautics.** 1959. Bureau of Naval Weapons, Navy Department, Washington, D. C. (NAVWEPS LS1, AD–436 719.) 74 pp.

4/65–21 **ECAC Thesaurus of Keywords.** September 1963. R. A. Lambert, J. E. Farrell and D. L. Flynn. 1st ed. Electromagnetic Compatibility Analysis Center, Annapolis, Md. (TM X042A1. AD–439 204.) 123 pp.; OTS price $10.00.

This thesaurus is the first edition in a continuing effort to provide coordinate-indexing access to literature in the fields of radio frequency interference (RFI), computer applications and mathematical modeling. This edition is strongest in RFI, but weak in the latter two areas. The thesaurus is divided into three sections: alphabetical list of keywords, equipment by AN/number, and equipment by various military and commercial designations. (author)

4/65–22 **Standardization Requirements for Technical Reports.** March 1963. David Taylor Model Basin, Washington, D. C. Rev. no. 2. Robert E. Lord. (DTMB 1010, PB 144 372. AD–601 706.) 44 pp.; OTS price $1.25.

The report presents the technical report standards established by the Bureau of Ships for its various laboratories and defines the different kinds of BuShips reports. It also outlines the procedures to be followed by the David Taylor Model Basin in the preparation and publication of technical reports consistent with BuShips policy. (author)

4/65–23 **Author's Guide for Technical Reporting.** July 1964. Alexander G. Hoshovsky. Office of Aerospace Research, Washington, D. C. (OAR 64 8. AD–605 443.) 33 pp.; OTS prices: HC $2.00/MF $.50.

This guide has been written to present some simple rules and suggestions for the authors of OAR sponsored scientific reports. Its contents are consistent with the majority of guides used by professional journals and with the current rules and directives of the Department of Defense and the United States Air Force. Emphasis is on titles, abstracts, and key wording, which are most important elements for indexing and retrieval. They are also key elements in a new Department of Defense form (DD Form 1473) which is now required as the last page in all technical documents.

4/65–24 **Recommendation for the Preparation of Indexes for Books, Periodicals, and Other Publications.** 1964. Report BS 3700:1964 of the British Standards Inst. 32 pp.; price 10s.

4/65–25 **Aquatic Biology Serials: Their Location and Characteristics.** August 1964. Charles W. Shilling and Mildred Benton. Biological Sciences Communication Project. George Washington University, Airlie Center, Washington, D. C. 109 pp.

A list is given of 1,627 serial publications in aquatic biology. An analysis is made of their country of publication, content characteristics, subject content, language, type of sponsorship, frequency of issue, birth and death, and abstract and index coverage. The location of these serials in one of 18 libraries is also noted.

4/65–26 **Journal Holdings of the National Documentation Centre.** List no. 1, May 1964. Thailand. National Research Council. 87 pp. In English and Thai.

4/65–27 **The U.S.S.R. and Eastern Europe: Periodicals in Western Languages.** 1964. 67 pp.; price $.45.

A bibliographic guide aimed at presenting an updated and somewhat enlarged body of information pertaining to the following countries: Albania, Estonia, Latvia, Lithuania, Bulgaria, Czechoslavakia, Hungary, Poland, Rumania, the Soviet Union, and Yugoslavia. The entries are arranged alphabetically by country and subject guide.

4/65–28 **Index to Scientific Journal Title Abbreviations Found in The Physical Review.** 1964. Phyllis A. Richmond, Report AIP/DRP 64–4 of the American Institute of Physics. New York. 143 pp.

Expansions for the 655 journal title abbreviations which could be identified have been provided in a form useful for finding the journals in a library. Differentiation among journals with like abbreviated citation forms was difficult. Accurate identification is given in the index by adding notes about dates, series and subsection of issuing body. All the different journals to which an abbreviation could refer are listed. (author)

4/65–29 **IDEP to DDC Report Number Cross-Reference Index.** March 25, 1964. Sylvania Electric Products, Inc., Waltham, Mass. (AD–435 650.) 38 pp.

This index is for use in determining the Defense Documentation Center Report Number when the IDEP Report Number is known. All IDEP reports announced in DDC's Technical Abstract Bulletins (TAB) up to and including February 1, 1964, are listed. (author)

4/65–30 **Coden for Periodical Titles. Supplement 1.** 1964. L. E. Kuentzel, Ed. ASTM Special Technical Publication No. 329–S 1. American Society for Testing and Materials, Philadelphia, Penna. (BATR). 166 pp.

An aid to the storage and retrieval of information and to communication involving journal references, prepared and maintained for Committee E–13 by the Wyandotte–ASTM Punched Card Project.

4/65–31 **Guide to Record Retention Requirement.** (Revised as of January 1, 1964.) 75 pp.; price $.44.

Designed to assist industry and the general public, this useful "Guide" is compiled from U.S. Statutes, and from regulations issued by the various Federal agencies. It contains 873 entries detailing the retention periods for the many types of records required to be kept under Federal laws and rules. The Guide tells the user what records must be kept, who must keep them, and how long they must be kept. Each entry also includes a reference to the full text of the basic law or regulation governing such retention. The booklet's index, numbering 1,724 items, lists for ready reference the categories of persons, companies, and products affected by Federal record retention requirements.

4/65–32 **Specifications for Library of Congress Microfilming.** 1964. Stephen R. Salmon. Library of Congress. Washington, D. C. 21 pp. USGPO price 25 cents. Catalog No. LC 1.6/4: M58.

This guide came about as a result of numerous inquiries concerning Library of Congress practices and recommendations in the field of library microfilming, and a need to specify the conditions under which microfilm would be considered for addition to the permanent collections of the Library. It draws extensively on the older *Guide to Microfilming Practices* prepared by the Committee on Photoduplication and Multiple Copying Methods of the American Library Association, and on the technical advice of Donald C. Holmes, Elmer S. King, and Charles G. LaHood. Applicable standards approved by the American Standards Association are cited throughout. (author)

4/65–33 **National Technical Information Services-World Wide Directory.** FID Publ. no. 359. May 1964. (Comp. by the Technical Information Service, National Research Council, Canada, on behalf of FID.) Ottawa. Copies may be purchased from FID, 7 Hofweg, The Hague, Netherlands, or from the Technical Information Service, National Research Council, Ottawa, Canada. 57 pp.; price $1.00 Can.

The guide is a listing of major channels in each country through which sources of technical information in any field can be located by enquirers unfamiliar with the country's technical facilities. Entries are listed by country, and include, when known, the scope of the facility described and working languages used.

4/65–34 **Foreign-Language and English Dictionaries in the Physical Sciences and Engineering. A Selected Bibliography, 1952 to 1963.** July 24, 1964. Tibor W. Marton. National Bureau of Standards Miscellaneous Publication 258. 189 pp.; price $1.25. (Order from the Superintendent of Documents, U.S. Government Printing Office, Washington, D. C. 20402, or from local U.S. Department of Commerce Field Offices.)

This bibliography, processed by punched-card, machine-sorting, and photo-offset techniques, lists over 2,800 unilingual, bilingual, and polyglot dictionaries, glossaries, and encyclopedias published during the past 12 years. The majority of the titles cited have English as the source or target language, or are English defining dictionaries. The entries are arranged in 49 subject classes; within each subject, the entries are listed alphabetically by language and within each language group, by author. Forty-seven foreign languages are represented in the compilation.

4/65–35 **National Microfilm Association Microfiche Standard Specification for Documents 8½ by 11 Inches and Smaller.** M–1–1963. October 1963. NAMN No. 66. pp. 42–52.

4/65–36 **The Documentation of Documentation. A Cumulative Index of Documentation Journals.** March 1964. Herbert Buntrock. NADO vol. 15, no. 1. pp. 53–56.

A cumulative index of 2,685 articles of five documentation periodicals (*Revue Internationale de la Documentation, American Documentation, Journal of Documentation, Nachrichten für Dokumentation*, and *Dokumentation*) is described. The structure of a thesaurus of documentation containing a systematic and an alphabetical list of descriptors (English, French, and German) with synonyms and definitions, is discussed. (*In German.*) Dokumentation der Dokumentation. Ein kumulativer Index von Dokumentations-Zeitschriften.

4/65–37 **Secondary Journals in Chemical and Biological Fields.** July 1964. R. F. J. McCandless, Elizabeth A. Skweir, and Maxwell Gordon. JCHD vol. 4, no. 3. pp. 147–153.

Briefly describes 38 secondary literature sources in the chemical and biological fields.

4/65–38 **Current Sources of National and International Pharmaceutical Market and Economic Information.** October 1964. W. A. Southern and Patricia J. Wilson. JCHD vol. 4, no. 4. pp. 237–244.

Characteristics of Information Resources

4/65–39 **The News Publication as a Medium of Scientific Communication.** January 1964. Philip L. Dopkowski. Biological Sciences Communication Project, The George Washington University, Washington, D. C. Communique #13. 33 pp. & 4 appendixes.

Seventy biology news publications serving primarily the scientific community were subjected to an analysis for content, format, audience, and source. Audience included the professional

scientist, graduate and undergraduate students, teachers, technicians, amâteurs, interested public, etc. Source included private, non-profit, academic, governmental and industrial organizations.

Of the 3,948 articles or items of information in the sample, 52% were found to have primarily human-organizational content, 48% research content. The leading categories represented in the greater number of publications were: Publication Announcements; Meetings, Conventions, Conferences; Organizational Progress; Personal Activities and Achievements; Promotions, Awards, Advancements; Current Investigations and Projects; Publication Reviews. Results of the format analysis indicate that the majority of news publications had qualities closely resembling those of the primary journal. 67% of the publications were found to serve primarily the professional biologist. 77% of the publications were sponsored by private, non-profit organizations. (author)

4/65–40 **Article-by-Article Coverage of Selected Abstracting Services.** October 31, 1964. Eugene Garfield and Irving H. Sher. Institute for Scientific Information, Philadelphia, Pa. NSF contract NSF C–332. 22 pp. and appendix.

The abstracting services selected for the study were *Biological Abstracts (BA)*, *Chemical Abstracts (CA)*, *Index Chemicus (IC), and Physics Abstracts (PA)*.—three major English language abstracting services and one smaller service for comparative purposes. Work involved scanning a total of 812,000 abstracts and keypunching cards for 219,000 abstracts of articles. Less total cards were keypunched than originally anticipated in the contract because of the much larger number of total abstracts that had to be examined before keypunching.

About 102,000 source items (from 613 journals) which later appeared in the 1961 *Science Citation Index* and about 8,000 additional items (from 45 journals from the fields of chemistry and physics), comprising a total of 109,705 articles published in 658 journals in 1961, were analyzed as to coverage by the four abstracting services. Of the abstracts keypunched, 99,682 were pertinent to the journals covered by the 1961 *Science Citation Index* files. Of these abstracts, 70,355 unique articles were involved—representing a redundancy factor of 1.42. Thus, 39,350 articles were not found to be abstracted by any of the four services.

A detailed analysis of errors from all sources has been made including checks by the services themselves. As a result, it is estimated that a total of about 770 unique items may have been erroneously assigned to the "unabstracted" category. This would still leave about 38,580 items not abstracted by any of the four services.

Of the 99,682 abstracts published, *Biological Abstracts* produced 33,899 with an average time lag of 31.2 weeks and median lag of 23.4 weeks; *Chemical Abstracts* produced 45,960 with an average lag of 26.8 weeks and a median lag of 22.3 weeks; *Index Chemicus* produced 8,080 with an average lag of 7.9 weeks and a median lag of 7.0 weeks; *Physics Abstracts* produced 11,654 with an average lag of 21.1 weeks and a median lag of 17.1 weeks. (authors)

4/65–41 **Characteristics of Scientific Journals, 1949-1959. Report of a Survey Conducted by the Office of Science Information Service.** 1964. 25 pp.; price $.30.

This report of a survey on the characteristics of scientific journals, Nat'l Science Foundation, represents an attempt to collect certain pertinent data on the primary research journals of the United States.

4/65–42 **Federal Department Libraries. A Summary Report of a Survey and a Conference.** 1963. Luther H. Evans and others. The Brookings Institution, Washington, D. C. 150 pp.

4/65–43 **American State Archives.** October 1964. Posner, Chicago University Press, Chicago, 416 pp.; price $7.50.

Mr. Posner traces the evolution of this country's archival institutions and reviews the present methods and functions of state agencies. Standards are established that may help to regulate techniques and to contribute to the full development of the archival profession.

4/65–44 **The Journal Literature of Physics. A Comprehensive Study Based on Physics Abstracts** (Science Abstracts, Section A) **1961 Issues.** 1964. Stella Keenan and Pauline Atherton. Report AIP/DRP PAI (1964) of the American Institute of Physics. New York.

A comprehensive study of the publication of physics literature has been made by analyzing the 1961 issues of *Physics Abstracts*. Information on 405 journals from 39 countries—a total of 20,287 articles—was collected. For each journal article, details about the year of publication, number of authors, language of article, type of abstract and placement in subject categories are now available in machine readable form. These data have been used to rank journals and countries in order of contribution to *Physics Abstracts* as a whole and to each subject field within physics. A profile for each country and for each journal, in terms of its coverage, is presented in tabular form. The distribution of the subject content of physics journal literature throughout the world is also shown. (authors)

4/65–45 **A Model Information Retrieval Network for Government, Science and Industry. A Proposed Basic Configuration for a National System of Interlinking Information Retrieval Networks.** May 1964. Jonker Business Machines, Inc., Gaithersburg, Md. (Contract AF49 638 1209. AFOSR 64 0942 AD–600 221.) 259 pp.; OTS price $16.50.

The report describes the technical, organizational and financial aspects of a model Information Retrieval Network which could be made operational at the present time. (author)

4/65–46 **Threat of Communist Technology: Proposals for Free World Cooperation in R&D and Information Transfer.** September 1, 1963. Hyman H. Album and Alexander G. Hoshovsky. Office of Aerospace Research, Washington, D. C. (OAR 18. AD–430 598.) 63 pp.; OTS price $6.60.

The scientific and technological trends in communist countries cannot help but evoke a sense of urgency. If these trends continue, the Free World must expect serious economic and political difficulties. To help prevent this, we have focused our attention on major problem areas—ineffective R&D cooperation of the Free World and its fragmented, cumbersome information system. As a partial solution we propose: a Free World alliance for science and technology; greater emphasis on education of scientists in countries emerging on the world scene; reduction of the one way flow of technical information to communist countries (or, inversely, increase our intake); a U.S. information system which would more effectively use government and private resources. (authors)

4/65–47 **The Information Center: Some Selected Examples.** August 4, 1964. Emory H. Holmes. System Development Corp., Santa Monica, Calif. (SP–1702. AD–606 174.) 34 pp.; OTS prices: HC $2.00/MF $.50.

The development of information centers, in contrast to traditional libraries, is discussed. Reasons for the failure of traditional libraries to meet special information needs of scientists, scholars, and technicians are pointed out. Information centers are discussed as a natural outgrowth of attempts to meet special information needs of scientists. Four selected operational information centers are discussed. Each center was selected to illustrate certain system design problems which should be considered when setting up an information center. (author)

4/65–48 **Why Are Information Centers Successful?** Alan M. Rees. *Parameters of Information Science.* pp. 173–176 (see 4/65–1).

It is argued that the information center is the visible incarnation of the invisible college in that it attempts to formalize the informal exchange of information amongst a closed set of users. Essential features are described in terms of the nature of the user groups, information needs, evaluation of input and output, search products, storage and retrieval techniques used and operating personnel. The success and vitality of the center is due to the closed and specialized nature of the user group rather than to the storage and retrieval systems used. (author)

4/65–49 **Documentation of Technical and Scientific Meetings.** Harry Baum. *Parameters of Information Science.* pp. 243–246 (see 4/65–1).

Practices in documentation of technical and scientific meetings cover the complete spectrum from no record to a complete record of all presentations and discussions in case-bound-book form. Part of the variation is accounted for by differences in the philosophical approach; part of it results from a lack of appreciation of documentation needs. The documentation problem will be presented as a result of the over-all philosophy, and recommendations for partial solution of the problem will be tendered. (author)

4/65–50 **Patents as a Source of Technical Information.** Girraj Mal. *Chemical Age of India.* vol. 15, no. 5, May 1964, pp. 585–588.

Describes patent literature, and lists patent organizations in 23 countries.

4/65–51 **Don't Overlook Patents.** Richard C. Willson, Jr. *Chemical Engineering,* vol. 71, February 3, 1964. pp. 79–84.

Describes how patent literature can be used to avoid duplication of research and as a springboard for new ideas. Legal aspects, reliability of patent literature, and search techniques are discussed.

4/65–52 **How to Get Information From Patents.** Richard C. Willson, Jr. *Chemical Engineering.* vol. 71, March 2, 1964. pp. 105–109.

Commercial indexing services which provide access to patent information are described. Methods for obtaining copies and summaries of U.S. patents are included. Part I of this two-part series appeared in the issue of Feb. 3, 1964, pp. 79–84.

4/65–53 **Should Science Information Be Centralized or Decentralized?** T. P. Kridler and G. S. Simpson, Jr. ISCT, no. 30, June 1964. pp. 100–104.

Compares U.S. and Soviet Scientific Information network. Describes the activities of the republic, regional and institute-level (ONTI) information organs, the All-Union Institute and Technical Information (VINITI), and the State Scientific-Research Institute of Scientific and Technical Information (GOSINTI).

4/65–54 **An Index to Ordnance Reports.** Alfred M. Anzalone and Gunther Cohn. JCHD vol. 4, no. 3, July 1964. pp. 169–172.

Discusses the data file set up by the Technical Information Section of Picatinny Arsenal, as a preliminary step to establishing an Initiator Information Center.

4/65–55 **An Evaluation of the Pesticide Literature—Problems, Sources, and Services.** Rhoda Ann Alverson. JCHD. vol. 4, no. 4, October 1964. pp. 204–208.

4/65–56 **Study of Physics Abstracts—Abstracting Three Typeso f oJurnals.** Pauline Atherton and Jean C. Yovich. JCHD 4(3):157–163. July 1964.

Surveys the time lag in coverage by *Physics Abstracts* of three types of journals: English language with no author abstracts; English language journals with author abstracts for articles but not for letters to the editor; and a foreign language journal with author abstracts in English or foreign language. The three types showed a degree of similarity in average time lag (3–4 mo.) and in rate of completion—an average of 5–6 mo. time lapse for 100% completion of a single journal issue. (BATR)

4/65–57 **Documentation in Electrochemistry.** T. M. Goerge and V. V. Parthasarathy. *Library Science,* vol. 1, no. 2, June 1964. pp. 147–183.

Studies the scatter and seepage of documents on electrochemistry among 592 periodicals. About 127 carry nearly 90% of the documents on the subject. Less than 4% are published in sources not dealing with electrochemistry or its penumbral areas. About 80% are articles. The highest number of citations in any one volume of the *Journal* is to the document published two years earlier. Approximately 40% of the citations in any one volume of the *Journal* are to documents published during the preceding five years, and 69% citations are to documents published during the preceding ten years. (author)

4/65–58 **Interlibrary Loan: A Reference Service.** Michael M. Reynolds. LIBT, vol. 12. January 1964. pp. 425–436.

Traces the history of lending among libraries, highlighting the problems of increased dependence on the large research library, the expanded use of research libraries by business and industry, and the needs of every community for books beyond the community's ability to provide.

4/65–59 **Structure of Information Facilities.** Sturktur von Informationseinrichtungen. Erich Pietsch. NADO, vol. 1, no. 1. March 1964. pp. 28–41.

Analyzes the structure of an information center and its behavior with regard to information, from the viewpoints of its preparation and of its use. Twelve theses underlying the practical structure of information are developed; definite American, West European, and Soviet examples are given for each thesis and special comments are made concerning the documentation of the Atomic Energy Commission in the U.S. and of the Institute for Scientific Information of the Academy of Sciences of the USSR (VINITI). (In German)

4/65–60 **Organisation of Science Information Activity in the Soviet Union.** A. I. Mikhailov. REDO, vol. 31, no. 4. November 1964. pp. 143–149.

Definition of the scope of science information and of information theory, now an autonomous discipline. Principle and organization of science information in the USSR. Activity of VINITI: abstracting, research, international relations, training. (In Russian, with a summary in English.)

4/65–61 **Processing Information for Research, Industry and the Public.** K. Rajagopalan. REDO, vol. 31, no. 4. November 1964. pp. 150–152.

Some problems of consolidation, assessment and dissemination of research information are presented and the need for organising national and international centres to feed current information to research workers and industry stressed.

4/65–62 **The Number of Review Articles in Various Subject Areas of Chemistry.** JCHD, vol. 3, no. 3. July 1963. pp. 139–141.

4/65–63 **Inter-Firm Cooperation in Technical Information.** W. T. Knox. *Research Management* vol. 7, no. 5, September 1964. pp. 337–348.

Three ways in which groups of industrial organizations have arrived at a satisfactory cooperative procedure for handling and disseminating pertinent scientific-technical information are described, namely, the central abstracting-indexing service of the Am. Petroleum Inst., the Internatl. Documentation Ring project of pharmaceutical manufacturers, and the subscription library and information services of the Inst. of the Aerospace Sciences in cooperation with NASA's Office of Scientific & Technical Information. (C.L.B.)

4/65–64 **The Information Goals of Engineers Joint Council.** Stanley Klein. SPLB, vol. 55, no. 3, March 1964, pp. 143–147.

Describes the E.J.C. philosophy and activities, including developing the thesaurus, coordinate-indexing training, and literature studies.

4/65–65 **Administrative Aspects of Interlibrary Loans.** William S. Budlington, SPLB, vol. 55, no. 4, April 1964. pp. 211–215.

Discusses staff functions, costs, and future considerations including photo-copies and special depository collections.

4/65–66 **Some Speculations on the Future of Interlibrary Loan.** George A. Schwegmann, Jr. SPLB, vol. 55, no. 4, April 1964. pp. 216–220.

4/65–67 **Ranganathan Series in Library Science. vol. X. Documentation and Its Facets.** 1963. S. R. Ranganathan, Ed. Asia Publishing House, New York. 639 pp.

The 70 chapters examine documentation, translation services, training in documentation, library resources, photo-reproduction, and documentation standards. The International Federation for Documentation (FID), the All-Union Institute for Scientific and Technical Information (VINITI), and the Indian National Scientific Documentation Centre (Insdoc) are described.

4/65–68 **Is the Scientific Paper a Fraud?** P. B. Medawar. *Tappi*, vol. 47, no. 7, 30A, 34A, 40A. July 1964.

Scientific-technical reports are considered fraudulent in the sense that they misrepresent the true thought processes involved in scientific discoveries, i.e., by falsely attributing the advances made to inductive reasoning. The inductive format of conventional scientific papers should be discarded. The "Discussion," traditionally the last part of a paper, should come first, followed by scientific facts and acts. Scientists should not be ashamed to admit that hypotheses (hunches) appear in their minds along uncharted by-ways of thought, that these are imaginative and inspirational in character, and that they are indeed adventures of the mind. (C.L.B.)

Manual Card Systems

4/65–69 **A Simple and Versatile Punch Card System for Bibliographic Use.** December 26, 1963. Gilbert A. Leveille, Howerde E. Sauberlich, and Wayne C. Goad. Army Medical Research and Nutrition Lab., Denver, Colo. Task 16. AMRNL 281. (AD–601 914.) 10 pp.; OTS price $.50.

A punch card system is described which should find application for general or specific bibliographic use. The system enables the individual to file a reference by subject, author and year. A reference can be cross-indexed in as many as four subject categories and up to 39,999 subjects can be indexed. The cards can be filed randomly and, consequently, the system can be maintained with a minimum of labor. (author)

4/65–70 **A System for Cataloging Reference Material.** June 7, 1955. R. P. Johnson, D. J. Blakeslee, and H. Skavdahl. RAND Corporation, Santa Monica, Calif. P–690. (AD–604 719.) 16 pp.; OTS prices: HC $1.00/MF $.50.

The needle-sort card system for cataloging reference material is discussed. By this system, reports are cataloged according to subject matter by notching out appropriate holes that line the four edges of a card. The sorting process is then accomplished by inserting a needle into the proper row of holes in a deck of such cards, thus permitting the notched cards to separate from the deck. This permits a quick examination of an itemized summary of all available reports that might be of interest.

4/65–71 **Organizing the Documentation of the State of Technique for Industry, Science, and Patent Practice.** Aufbau einer Dokumentation des Standes der Technik für Wirtschaft, Wissenschaft und Patentwesen. Herbert Danner, NADO, vol. 15, no. 2, June 1964. pp. 74–81.

In large fields of knowledge with over 1,000 total characteristics, surface punch-cards are most suitably used. When technically possible, a separate card file should be set up for each field complex. The analysis of the references should be limited to the important characteristics, and yet be accurate. In German.

4/65–72 **Records Retrieval.** *Transportation & Distribution Management*, vol. 4, no. 3. 1964. pp. 37–40.

Kodak's traffic department uses microfilm system to save 2,800 man-hours yearly, eliminate 80 file cabinets. Describes how Eastman Kodak tied a record microfilming operation into their punch card freight payment control system—with spectacular results.

Reports of Facilities in Operation – Libraries

4/65–73 **Machine Processing in a Special Library.** Vol. 1. A Survey of Case Studies. July 15, 1964. Louise Schultz. Report SP–1679/001/00 of the System Development Corporation, Santa Monica, California. (AD–446 610.) 43 pp.

The range of application of EAM and EDP techniques to the technical operations and service mission of the special library is typified by several summaries. A table of principal features of each system studied aids comparison of the applications. Source literature on each system is cited. (author)

4/65–74 **Index Preparation and Library Processing at Monsanto Chemical Company's Research Center.** 1964. IBM Corp. Data Professing Div., White Plains, N. Y. Application Brief K20–0004–0.

4/65–75 **Establishment of a Coordinate Indexing Retrieval System for the Air Force Materials Laboratory.** Final summary rept., December 1, 1960 - November 30, 1963. Edward A. Janning. Dayton U. Research Inst., Ohio. November 1963. (Contract AF33 616 7698. Proj. 7381. Task 738105. RTD TDR63 4263. AD–428 423.) 71 pp.; OTS price $7.60.

A coordinate indexing system for a library containing over 10,000 documents has been established. The primary feature of this system is the ability to conduct either manual or computer searches. Manual searches are made using a dual dictionary that has been prepared by the University's NCR 304 computer. A thesaurus of over 18,000 multidisciplinary terms for this system has also been prepared. This report discusses in detail the indexing of documents using deep indexing with links and roles, the procedure used in editing the generated vocabulary, and the various machine operations used in mechanizing the system. Briefly discussed are various studies made, or in the planning stage, to evaluate several aspects of the system; in particular, an evaluation of the use of links and roles in the retrieval of information. (author)

4/65–76 **The Automated Approach to Technical Information Retrieval. Library Applications.** March 1964. John J. Nicolaus. Bureau of Ships, Washington, D. C. Rept. no. 250 210 2. (AD–446 643.) 44 pp.

This publication describes the Information Storage and Retrieval System, Project SHARP (SHips Analysis and Retrieval Project) developed for the Bureau of Ships Technical Library in a cooperative project between the Bureau Library and the David Taylor Model Basin. Documents can be retrieved using bibliographic data, subject matter and combinations of both. The Engineers Joint Council-advocated system of Roles and Links has been utilized as the method of indexing, and has been implemented as the means of search on the IBM 7090 and 1401 computer facilities at the David Taylor Model Basin for the Technical Library. Other library in-house applications are described, and additional features planned outlined. (author)

(For reference only at each of the DDC offices. This report cannot be satisfactorily reproduced. DDC does not furnish copies; available from GPO, Washington 25, D. C. at $.50)

4/65–77 **Automation of the Penn State University Acquisitions Department.** Thomas Minder and Gerald Lazerick. *ADI Proc. Pt. 3* (see 4/65–2).

A work flow and cost study of the Acquisitions Department of Penn State University Library was made. This served as a basis for the design of a new ordering system using a digital computer. (authors)

4/65–78 **DSD Technical Information Center—A Total Approach to Library Mechanization.** H. S. White. *ADI Proc. Pt. 3* (see 4/65–2).

The IBM Data Systems Division Technical Information Center (TIC) provides an operating developmental system for integrated and compatible mechanized processing of technical

information received within the organization. The system offers several advantages: (a) it is a sophisticated mechanized system for dissemination and retrieval; (b) it is compatible with all library mechanized records produced under a standard processing format within IBM libraries, providing such traditional tools as 3 x 5 catalog cards, circulation records, and overdue notices; (c) it is reversible, so that discontinuation of machine processing would not cause gaps in the library's manual records; (d) it is controlled, producing statistical evaluations of its own program efficiency; (e) it is user-oriented, providing 24-hour copy access and immediate microfilm access to its documents; (f) it is relatively simple, relying on the IBM 1401 Data Processing System for all its processing and output. Since the system has been operating for over a year, the conclusions drawn are based on actual experience. (author)

4/65–79 **Utilizing KWIC for Effective Procedures Administration.** N. Demain and J. S. Reed. *Parameters of Information Science*, pp. 117–119. (see 4/65–1)

A system is decribed wherein the Company's published policies and procedures are treated as a collection of technical documents. A master file is created which includes for each procedure a title, additional keywords, procedure number and classification code. An existing IBM 1401 KWIC program was modified to provide a keyword list with cross reference ("see" and "see also") statements and a bibliography section in the form of a classified index. System outputs include keyword and classified index for each manual and selected analytical reports by subject to insure adequate coverage and minimize redundant effort. (authors)

4/65–80 **Integrated Library Management Systems Concept.** B. E. Lamkin. *Parameters of Information Science*, pp. 141–147. (see 4/65–1)

The systems concept which we developed (the Library Management System) uses computers to perform normal library operations. It consists of seven basic operating units, each of which may be run in combination or independently. Specifically, the system includes orders, receipts, circulation, periodicals control, service. index, and reporting. Examples of outputs generated by the operating units are given and discussed. The historical background and some of the techniques used in operating the system are reviewed. (author)

4/65–81 **Experience with EDP Support of Individual's File Maintenance.** Everett M. Wallace. *Parameters of Information Science*, pp. 259–261. (see 4/65–1); also available as report SP-1646 of the System Development Corp., Santa Monica, Calif. (AD-443 751.)

During the past two years, SDC has been experimenting with various tools for supplying individuals with improved means of maintaining personal files and special records. We have previously reported on a prototype system at the 1963 ADI meeting. This paper reports the experience of users of the system and on modifications and improvements adopted in the light of that experience. (author)

4/65–82 **The M.I.T. Technical Information Project—A Prototype System.** M. M. Kessler, E. L. Ivie, W. D. Mathews. *Parameters of Information Science*, pp. 263–268 (see 4/65–1).

A working model of a technical information system has been designed and constructed. The system will be put to experimental use in a realistic test environment. The library now consists of 35.000 papers from nineteen physics journals kept up-to-date. A number of computer programs have been written that allow a flexible manipulation of the recorded elements of the literature for purposes of retrieval, study and testing. A unique feature of the system is its use of a remote, time-sharing computer console that allows real-time interaction between a user and the literature. (authors)

4/65–83 **Upgrading a Company's Information Systems.** John E. Ryan. *Automation*, vol. 11, May 1964. pp. 48–52.

Describes the origin. services, and results of the Internal Automation Operation at General Electric Co., Schenectady, N. Y.

4/65–84 **Mechanization of Library Procedures in the Medium-sized Medical Library: II. Circulation Records.** Irwin H. Pizer, Isabelle T. Anderson, and Estelle Broadman. BMLA, 52(2), April 1964. pp. 370–385.

Utilizing equipment and supplies for a Bookamatic circulation control system already available, and IBM equipment in the Computer Center of Washington University, a system has been developed which permits faster access to information needed for locating books in circulation, and for processing overdues on an end-of-semester basis without major disruption of files. Plastic borrowers' cards are assigned to faculty, staff, and students: faculty and staff cards bear University payroll numbers, and students are numbered serially by class.

4/65–85 **The MEDLARS System.** Charles J. Austin. DTMN, vol. 10, no. 12. December 1964. pp. 28–31.

4/65–86 **Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries.** Robert R. Freeman, et al., JCHD, vol. 4, no. 2, April 1964. pp. 107–112.

Experience in the use of computers for searching Chemical Titles over a period of more than a year is reviewed. The results of experiments have provided a current-literature alerting system at Eli Lilly and Co., a current awareness system and a retrospective search file at Olin Mathieson Chemical Corp., and indexed annual bibliographies for various areas of chemistry, at the Chemical Abstracts Service.

4/65–87 **Automation and Library Systems.** Theodore Stein, LIBJ, vol. 89, no. 13, July 1964. pp. 723–734.

Reports an investigation which was essentially a literature search and analysis of published and unpublished material dealing with the application of machines to library problems. This was supplemented by visits to libraries where data-processing equipment is used. Four broad types of application are discussed: 1. direct machine search of catalogue files (machine information retrieval); 2. machine indexing and abstracting; 3. routine record keeping; 4. catalogue maintenance and preparation of book catalogues.

1. The systems covered are mainly those employing electronic computers as only these appear to have the capacity for application to a large reference library. There is frequent indication that present techniques in machine searching do not adequately substitute for human judgement. Machine searching supplemented with human judgement is often described.

2. Using machine analysis to assign subject headings or prepare abstracts is at the development stage. At present the results are inferior to those produced using conventional methods.

3. Data-processing techniques for routine record keeping, such as book ordering, circulation control, serial control, and accounting, can be successfully applied. Required operations can be completely defined beforehand, the desired results are fairly constant and completely known beforehand, and the data to be operated needs no interpretation.

4. If a library catalogue is transcribed to some machine-readable form the following outputs may be produced: a book catalogue of the entire collection; various supplementary lists, namely, machine-prepared bibliographies, new acquisition lists, unions lists in library systems, special-purpose lists; statistics on the collections. (JDOC)

4/65–88 **New Technique Provides Means for Full Utilization of Technical Information.** John J. Nicolaus, NMRV, vol. 9, no. 6, June 1964, pp. 29–30, 32.

4/65–89 **The National Library of Medicine and MEDLARS.** REDO Scott Adams and Taine Seymour, vol. 31, no. 3, August 1964, pp. 107–110.

The MEDLARS system, representing computerization of the indexing performed by the National Library of Medicine for over 80 years, is now becoming operational. The January 1964 issue of the *Index Medicus* has been successfully prepared and composed by the computer. Future issues will be composed by a high-quality, high speed, photo-composing device known as GRACE. Based on this central resource, the Library will be

offering recurring bibliographic services to national scientific and professional groups as well as to government agencies, and will be providing a demand search service nationally in answer to sophisticated inquiries. (authors)

4/65–90 **Use of Mechanized Equipment in the Production of Library Records for Manual Handling or Computer Manipulation.** Herbert S. White, STNW, vol. 18, no. 2. Summer 1964. pp. 23–26.

4/65–91 **Automated Circulation at a Government R&D Installation.** I. Haznedari and H. Voos. SPLB, vol. 55, no. 2. February 1964. pp. 77–81.

Describes the circulation procedures for books and reports, using IBM 026, 082, 1401, and 1403 at the Technical Information Section, Picatinny Arsenal.

4/65–92 **PIL (Processing Information List) or a Computer-Controlled Processing Record.** Roger P. Bristol. SPLB, vol. 55, no. 2, February 1964. pp. 82–86.

The processing record system at the Thomas J. Watson Research Library, IBM, indicates what is on order, what has been received, where it is in the processing cycle, and how long it has been there. The system also gives as machine produced by-products: the shelf-list card, the circulation card, and the book pocket label.

4/65–93 **Practical Application of Automation in a Scientific Information Center—A Case Study.** C. K. Bauer. SPLB, vol. 55, no. 3, March 1964. pp. 137–142.

This case study concerns Lockheed-Georgia Company's Scientific and Technical Information Department. The computer-processed outputs are described and cost data are included.

4/65–94 **Preparationo f a Book Catalog.** Erik Bromberg, G. A. Dubinski, and Donn Remington. SPLB, vol. 55, no. 9, November 1964. pp. 611–614.

Describes the project at the Library of the U.S. Department of the Interior in Portland, Oregon. Gives examples of catalog and cost figures.

4/65–95 **KWIC and Easy? A Librarian's View of a Computer-Based Technical Reports Announcement System.** Paula M. Strain, SPLB, vol. 55, no. 9, November 1964. pp. 614–618.

Describes the KWIC index produced at IBM's Space Guidance Center.

4/65–96 **Interim Report on Library/USA.** Joseph Becker. *Stechert-Hafner Book News,* vol. 19, no. 1, pp. 1–3. September 1964.

The "Library/USA" exhibit at the N. Y. World's Fair, sponsored by ALA, SLA, and ADI, is described and illustrated. It comprises a ready reference service for visitors, a computer and communications system demonstrating the impact of technology on libraries, a display of recent developments in library service to children, and an adult reading-browsing area of 2,000 books, reflecting selections made by J. T. Babb's committee of librarians for the White House library. (C.L.B.)

4/65–97 **A Book Catalog at Work.** Margaret C. Brown. LRTS, vol. 8, no. 4. Fall 1964. pp. 349–358.

Since the fall of 1963, the Free Library of Philadelphia has had a book catalog representing all items added after January 1, 1961 to the collections of the Library's Extension Division. The Library's experience is discussed in terms of 5 questions: Why do you do it? How do you do it? How much does it cost? Has conformity to the requirements of the machine changed the kind of cataloging you do? How do readers like it, and has your staff accepted it?

Before deciding to use the Compos-O-Line camera to produce the book catalog, 2 other sequential card cameras (Kodak's List-O-Matic and VariTyper Corporation's Foto-List) and the IBM printer were studied. Compos-O-Line was selected because it could accept 3 x 5 cards, and most of the catalog cards already prepared would be usable. Since catalog cards are still being multilithed for the main catalog, extra cards can be multilithed for the so-called "master file," and the book catalog can thus be reproduced as a kind of by-product of the general cataloging operation. (author)

4/65–98 **Book Catalogs as Supplements to Card Catalogs.** Phyllis A. Richmond. LRTS, vol. 8, no. 4. Ball 1964. pp. 359–365. pp. 359–365.

Using IBM data processing equipment, supplementary single-line-entry book catalogs have been made for the science libraries of the University of Rochester, extending the usage of the conventional card catalog. A program SHORTTITLE for the IBM 7074 computer has been developed to compile a unified catalog for several libraries. Similar catalogs can be made by copying pertinent sections of any shelf list. (author)

4/65–99 **The Compact Book Catalog—by Photographic Process.** Bob Jones. LRTS, vol. 8, no. 4. Fall 1964. pp. 366–369.

The newly formed Junior College District in St. Louis, with a central instructional resources center and three campuses produced book catalogs by Compos-O-List camera, the variable aperture adjusted to four different bites. Cumulative supplements were issued, with a total cumulation planned each year. This same type book can be produced by the Photo-Director or Itek camera, both requiring that the cards be shingled in pages rather than photographed individually. For the first yearly cumulation, a new method developed by Alanar Book Processing Center was used. By this process, the whole LC card was photographed in page layout, the pages being numbered and having a letter designation for each card. From this, punched cards and computer tape (1401) were made, run off by author-title and by subject, photo-reduced for clarity, and hard bound. Each entry in the author-title and subject indexes contains author, title, date, the editor or edition if needed, the classification and cutter numbers, and a code number referring back to the book containing the full LC card. A code number such as 70R refers to page *70,* space *R* in the book catalog if the full entry is needed; otherwise, the book can be located on the shelves with the classification number. Plans for subsequent issues are indicated. (author)

4/65–100 **The Computerized Catalog: Possible, Feasible, Desirable?** Wesley Simonton. LRTS, vol. 8, no. 4. Fall 1964. pp. 399–407.

An attempt to relate the general parameters of computer operations to the accepted goals and procedures of present cataloging operations, considering whether these goals and procedures can, should, or must be modified to take advantage of computer techniques and capabilities. Concludes that a computer-produced book catalog is certainly *possible;* and *desirable* if the filing problem can be solved, eliminating possible human errors. *Feasibility* remains a question, at least if we consider cost. Book catalogs are being produced by computers today, but for the most part in relatively small, technical libraries with ready access to a computer. There is great need for experimentation, with close attention to cost, both of present and proposed methods, in a wide variety of types and sizes of libraries, before the question of feasibility can be answered. (author)

Reports of Facilities in Operation – Other Information Centers

4/65–101 **A Retrieval System for Searching Defense Documentation Center Magnetic Tape Files.** October 1963. K. L. Kelley and R. C. Anderson. Aerospace Corp., Los Angeles, Calif. Rept. no. ATN 63 9990 5 (AD–427 709).

A document retrieval system was developed with the cooperation of the Defense Documentation Center. The Aerospace Corporation has converted copies of the DDC retrieval tapes to a format acceptable by an IBM computer and developed a retrieval computer program. This report describes the logic of the computer program developed, and provides instructions for use of the program. (author)

4/65–102 **Information Retrieval in the Personnel Department. A Survey of Methods Used in Scientific and Engineering Organizations.** April 1964. Robert A. Dickmann, Olinda Elliott, and Leslie A. Hubbard. Applied Physics Lab., Johns Hopkins University, Silver Spring, Md. TG574. Contract NOw62 0604c (AD–601 488). 31 pp.; OTS price $1.00.

Prepared in cooperation with George Washington Univ., Washington, D. C., the present study was undertaken to determine the approach scientific and engineering organizations are taking to automate personnel records, the nature of such automated systems, the degree of success which has been encountered, and the efforts made toward developing a structure of technical skills. Questionnaires on automation and information retrieval in the personnel department were sent to 256 organizations. One hundred seven firms responded, 95 with completed questionnaires. After the results were summarized, the following conclusions were drawn: 1. An actual need for automation of personnel records was first seen as a firm approached 500 professional employees. 2. As continued growth occurred (up to 1,000 professional employees), the need for automation was confirmed. 3. After clerical problems of preparing statistical reports were solved, in the 100 to 1,500 groups, firms began to load more detailed information about each employee into the system so that comprehensive lists of employees could be provided. 4. In the 1,500 to 2,000 group, all of the responding firms relied on automated records to some extent, even though the sample of firms in this category was very small. 5. In firms with 2,000 or more professional employees, there was a trend to integrate several automated subsystems of personnel records. (author)

4/65–103 **USAF Mechanical Properties Data Center.** Belfour Engineering Co., Suttons Bay, Mich. AF33(657)–11218 Progress Report 2, R. C. Braden and C. S. Wright. October 15, 1963. AD–422 296, 15 pp. Progress Report 2, R. C. Braden. May 15, 1964. AD–440 383. 26 pp. Progress Report 3, R. C. Braden. August 15, 1964. AD–444 486.

These reports cover the continued operation of the U.S.A.F. Mechanical Properties Data Center, including the expansion of data files, capabilities, and data output. This effort is under the direction of the Research and Technology Division, Air Force Materials Laboratory, Materials Application Division, Materials Information Branch, Wright-Patterson Air Force Base, Ohio. (author)

4/65–104 **The Mechanical Properties Data Center Operation and Expansion.** Summary technical rept. for April 1, 1963–March 31, 1964. R. C. Braden and C. S. Wright. Belfour Engineering Co. Suttons Bay, Mich. August 3, 1964. Contracts AF33 615 1061, AF33 657 9149, Proj. 7381, Task 738103. ML TDR64 235 (AD–605 815). 105 pp. OTS prices: HC $4.00/MF $.75.

The report discusses the content, use and operation of the Mechanical Properties Data Center. Search frequency, data input, document acquisition, and link-role indexing are topics of the report. A review of project activities expressed as percentages of expended effort is presented. (author)

4/65–105 **STARSHINE Program for Storage, Retrieval, and Analysis of Shock—Damage Information.** December 1963. Ronald E. Baker and Natalie T. Goldberg. David Taylor Model Basin, Washington, D. C. Proj. SF013 10 02, Task 1799 DTMB 1782 (AD–433 795). 19 pp.; OTS price $1.60.

As the result of a program of shock hardening and underwater explosion shock tests on operating ships authorized by the Chief of Naval Operations, large quantities of information are being accumulated on shock damage to and malfunctioning of equipment, and on the circumstances under which these occur. A system for storing, collating, and analyzing these data is essential to their effective utilization to improve the Fleet by increasing the reliability of vital equipment and to evaluate the effectiveness of the shock-hardening program itself. Program STARSHINE, which utilizes the LARC computer, was devised to meet this need. (author)

4/65–106 **Mechanization of the UDC. Final Report on Pilot Project to Further Explore Possibilities for Mechanization of Universal Decimal Classification (UDC) Schedules.** Malcolm Rigby. June 3, 1964. American Meteorological Society. Meteorological and Geoastrophysical Abstracts. Washington, D. C. 50 pp.

A pilot project for the purpose of determining the feasibility, costs and problems involved in mechanical storage, retrieval and printout of UDC schedules in any or all languages, especially in the three official languages (English, French and German) is described. The factors considered were: 1) the procurement of schedules; 2) preparation of copy for punching (including transliteration of Russian and Romanization of Japanese characters); 3) punching and verifying cards; 4) editing (proof) computer run; 5) correcting card file; 6) taping cards; 7) programming for single language, for multilingual presentations and for updating; 8) final computer runs for study or reproduction; 9)g reproduction problems; 10) cost analysis for limited sample, for a larger sample or for entire body of UDC schedules, including extensions and corrections.

The sample treated, involving 40,000 punched cards, included about 27,000 headings from 3 full editions (English, French and German), 1 intermediate edition (Russian), 4 special editions, and the abridged editions in 12 other languages, using only class 52 (Astronomy and Astrophysics) and class 55 (Geology-Geophysics). The average cost for punching and verifying was as predicted, regardless of language, about $91 per 1,000 cards. (author)

4/65–107 **The M.I.T. Technical Information Project. I. System Description.** November 1964. M. M. Kessler, Massachusetts Institute of Technology, Cambridge, Massachusetts.

A working model of a technical information system has been designed and constructed. The working literature is taken from 21 journals in the field of physics. The system utilizes remote consoles to access a time sharing computer facility (Project MAC). Programs have been developed for a large variety of search and processing techniques in real time as well as for delayed output. The system is intended to be a prototype operating in a realistic test environment. (author)

4/65–108 **Electrical and Electronic Properties of Materials: Information Retrieval Program.** Rept. for January 22, 1963-January 31, 1964. Thayne Johnson and Dana H. Johnson. Hughes Aircraft Co., Culver City, Calif. April 1964. Contract AF33 616 8438, Proj. 7381, Task 738103, ASD TDR62 539 P3 (AD–602 411). 89 pp.; OTS prices: HC $3.00/MF $.75.

The use of the Electronic Properties Information Center's collection, a modified pre-coordinated index, indexed for storage and retrieval either manually or through a 1401 Computer, is treated briefly and note is made of changes. The nine categories of materials searched, indexed and documented are semiconductors, insulators, ferrites, ferroelectrics, ferromagnetics, electroluminescents, thermionic emitters, metals and superconductors. The overlapping and interdependence of materials in these categories is discussed. Approximately 11,500 articles had been collected and indexed and 1,200 pages of data sheets published as this report was completed. (author)

4/65–109 **Applied Research Program Aerospace Intelligence Data System (AIDS). Volume I.** Quarterly rept. no. 3 for period ending February 28, 1962. Standard Electrical Products Co., Dayton, Ohio. Contract AF19 626 10. (AD–428 642.) 27 pp. OTS price $2.60.

4/65–110 **Applied Research Program Aerospace Intelligence Data System (AIDS). Volume II. Multiple and Simultaneous Services from a Single Computer Complex.** Quarterly rept. no. 3 for period ending February 28, 1962. 1 vol. IBM Research Center, Yorktown Heights, N. Y. Contract AF19 626 10 (AD–428 306). OTS price $10.10.

4/65–111 **SABIR2 (Semi-Automatic Bibliographic Information Retrieval, Second Version) Documentation. Volume I.** April 1964. Carol S. Haworth. Naval Postgraduate School, Monterey, Calif. NPS, M RP44 V1. (AD–603 316). 174 pp.; OTS prices: HC $5.00/MF $1.00.

SABIR2 facilitates an important service now available to patrons of the U.S. Naval Postgraduate School Library. A sophisticated machine information storage and retrieval system provides one of the few feasible solutions to a problem that plagues scientific communities throughout the United States. The problem can best be described as the determination of the method by which information on a specific subject or a group of subjects can be most easily, accurately, and expeditiously obtained.

4/65–112 **SABIR2 (Semi-Automatic Bibliographic Information Retrieval, Second Version) Documentation. Volume II.** April 1964. Carol S. Haworth. Naval Postgraduate School, Monterey, Calif. NPS, M RP44 V2. (AD–603 317). 398 pp.; OTS prices: HC $7.00/MF $1.75.

The report includes the SABIR2 program and is a continuation of AD–603 316.

4/65–113 **PALL: Rand's Automated Address Book.** September 1964. G. E. Bryan. RAND Corp., Santa Monica, Calif. RM–4258 (AD–605 456). 89 pp.; OTS prices: HC $3.00/MF $.75.

The PALL program (Prints Address Lists and Labels) has been in use since early 1963 to produce the various pieces of paper—distribution lists, receipts, mailing labels, and order forms —required in the distribution of RAND publications. It maintains a master file of some 850 organizations. In design and operation, the program resembles a two-pass assembler and makes extensive use of modern string-handling, pointer, and dictionary techniques. Section I of the Memorandum is a brief discussion of the origin, operation, and advantages of PALL. Section II contains the instructions for users of the program; Section III, the instructions for operators. Section IV describes the program's operation, step by step, and Section V discusses its formats. (author)

4/65–114 **RAPID, A System for Retrieval through Automated Publication and Information Digest.** June 1, 1964. Louise Schultz. Rept. no. SP 1678 000 00. System Development Corp., Santa Monica, Calif. (AD–443 774). 15 pp.

This paper describes a conceptual model for an advanced information system that can improve dissemination without increasing storage problems. Referred to as RAPID (Retrieval through Automated Publication and Information Digest), the concept comprises a coordinated, semiautomated dissemination and retrieval system built around a special science newspaper. The RAPID is described in terms of its applicability to a technical and management community of broad subject base, large size, and geographic dispersion. A comprehensive, but preliminary, system diagram is provided. (author)

4/65–115 **KAPL's Automated Information Retrieval System.** C. J. Schmidt, R. P. Wesley, and C. H. Hunter. *Parameters of Information Science.* pp. 177–184. (see 4/65–1)

This paper presents a status report on the automated IR system in operation at the Knolls Atomic Power Laboratory (KAPL). Significant system features including services provided, subject classification, search logic and strategy, and supporting statistical and maintenance routines are described. Adoption of the microfiche concept as an integral part of the IR system to provide an active document reproduction and dissemination medium is discussed. Economic considerations and system design aspects considered important to successful control and operation are also presented. (authors)

4/65–116 **Agencywide Information Retrieval for FDA: A Beginning.** Alfred Weissberg and Irving M. Gardoff. *Parameters of Information Science.* pp. 185–189. (see 4/65–1)

The Food and Drug Administration has completed the preliminary design phase of an integrated, agencywide information system for its scientific and program activities, and is beginning the detailed design of the system. The system will be based upon a computer-operated central reference file containing canonical lists, synonyms, and references to the files holding the information. These subject files would be located in physical proximity to their heaviest users and could be approached independently of the central index. This report discusses this system concept in relation to the day-to-day responsibilities of the agency. (authors)

4/65–117 **SHIRTDIF—A System for the Storage, Handling and Retrieval of Technical Data in Image Format.** Andries van Dam and David Evans. *Parameters of Information Science.* pp. 323–329. (see 4/65–1)

This paper describes the design of an integrated system of storing, handling, and retrieving information in image format. A digital computer is used to handle indexing, and yields either a description of the document called for, or the location in the analog memory in which the document is stored. The system is based on National Cash Register's photochromic micro-image bulk storage used in conjunction with a list processing program for indexing and retrieval. Specific advantages of the system taken as a whole are outlined. (authors)

4/65–118 **SCAN—A Flexible Scientific Literature System.** A. W. Elias. *Parameters of Information Science.* pp. 121–129. (see 4/65–1)

A case study is presented which resulted in the Wyeth Laboratories literature abstracting, publication and retrieval system. The flexibility inherent in simple unit record approaches to literature handling is shown with particular emphasis on the IBM 870 Document Writing System. (author)

4/65–119 **Data Retrieval.** D. B. Lloyd. CMPJ. (Gr. Brit.) vol. 7, no. 2. July 1964, pp. 110–113.

Any filing system storing important information is bound to have questions asked of it. A magnetic-tape filing system is no exception, and when some question is posed it is essential that a precise answer can be quickly and economically obtained. A General Information Retrieval Program has been written to meet this need, and has been extensively used for the last eighteen months with considerable success. (author)

4/65–120 **Systems for Information Retrieval in Small Laboratories.** Paul Khan and Jocelyn Rosen. *Food Technology,* vol. 18, June 1964, pp. 56–59.

Describes the information center in the Continental Baking Company's Research Laboratories, to illustrate the review of philosophy, systems, and tools.

4/65–121 **A Punched Card Retrieval System for Automobile Accident Cases.** John L. Garland. MULL. September 1963. pp. 130–134.

The author describes an elementary system for indexing state and federal court decisions dealing with automobile collisions, and for retrieving pertinent citations on the basis of fact elements, law elements, bibliographical elements, or any combination of the three.

4/65–122 **SAAMA Tests New Procurement Concept.** NAMN, no. 71. August 1964, pp. 21–28.

Aperture card system for engineering data.

4/65–123 **Automation Comes to the Real Estate World.** Robert M. Austin, SSTM, vol. 5, no. 6, November 1964, pp. 28–29, 36.

(System can sort more than 1,000 listings to find properties suited to exact needs of each buyer.)

Reports of Proposed Facilities – Libraries

4/65–124 **Automated Routines in Technical Services.** February 1964. Patricia T. Sievers and Paul J. Fasana. Air Force Cambridge Research Labs., Bedford, Mass. AFCRL 64 70 (AD–435 615). 21 pp. OTS price $1.60.

This paper discusses the approach to automation taken by the AFCRL Research Library, applicable to libraries in general.

A brief review of the current market of equipments and techniques is followed by guidelines this library will use in order to effect total systems automation. The AFCRL Research Library monograph processing system is discussed as an application of the techniques described. It covers the preliminary search routines using an MP-3 Polaroid camera, the cataloging routines making use of various machine-generated authority lists, the bibliographic encoding and catalog-card generation using tape typewriters, and the special-purpose data-processing device, the Itek Crossfiler, developed for this library. (author)

4/65–125 **EDP for Reclassification and Technical Services in the Los Angeles State College Library.** February 4, 1964. Joel M. Kibbee. System Development Corp., Santa Monica, Calif. Rept. no. TM 1731 000 00 (AD–431 014). 74 pp. OTS price $7.60.

System Development Corporation, under contract with Los Angeles State College, undertook a joint study with the Library Staff to investigate: 1. using EDP equipment for any of the clerical procedures necessitated by reclassification of books classified by the Dewey Decimal System; 2. using any EDP procedures developed for reclassification in handling new acquisition; and 3. the possibility of converting from a card to a book catalog. (author)

4/65–126 **The Automated Approach to Technical Information Retrieval, Library Applications.** 1964. 44 pp. price $.30.

Describes in some detail measures the Bureau of Ships has taken in an attempt to resolve some of the attending problems inherent in the handling, storage, and retrieval of scientific and technical information literature in their Technical Library.

4/65–127 **CARES—A Proposed Cartographic Retrieval Service.** Joseph C. Donohue and Charles P. Hedges. *Parameters of Information Science.* pp. 137–140. (see 4/65–1)

Information content of maps and charts is increasingly important to decisions involving military security and development of economic resources, including hydroculture. Present methods for identification, analysis and retrieval of maps are inadequate. A central agency is suggested, not to collect maps but to identify maps, and to refer researchers to sources, archival and trade. Problems of administration, classification and analysis are identified, along with possible solutions, including EDP use. (authors)

4/65–128 **Data Processing Techniques in an Information System.** L. Miller, S. Rosen, and R. Swid. Parameters of Information Science. pp. 149–153. (see 4/65–1)

Traditionally, information systems have been designed around elaborate theories of information, storage and retrieval. This unilateral approach tends to neglect sound data processing techniques, i.e. modular programming, powerful and flexible file formats and the ability to integrate new requirements at minimal programming cost. This is an unwieldy and expensive practice in a production environment. The operating system described herein tempers theory with practical concepts. Topical scope includes file organization, executive program control, semiautomatic indexing, and file searching. (authors)

4/65–129 **RAPID, A System for Retrieval through Automated Publication and Information Digest.** Louise Schultz. *Parameters of Information Science.* pp. 79–87. (see 4/65–1, also 4/65–114)

4/65–130 **A Proposed Large-Scale Information and Retrieval System.** Julius Frome. *Parameters of Information Science.* pp. 131–136. (see 4/65–1).

The system is discussed with emphasis on improvements through such features as: 1. the use of a free language to provide flexibility and ease in searching and indexing; 2. roles and links; 3. descriptor weighting by both searcher and indexer; 4. hierarchical search capabilities; 5. referral services (scientists' capabilities); 6. selective dissemination of abstracts; 7. literature searching with random access storage devices; and 8. ability to search ranges of numerical data for retrieval purposes. (author)

Reports of Proposed Facilities — Other Information Centers

4/65–131 **U.S. Army Engineering Data and Information System (EDIS) Concept and Action Plan Report.** July 1964. Stanley A. Goldberg, Roland A. Ham, Ralph E. Armbruster, Maurice E. Taylor and Walter Govinksy. Army Research Office, Office of the Chief, Research and Development, Washington, D. C. (ARO EDIS1. AD–444 700L.) 37 pp.

(Notice: All release of this document is controlled. All certified requesters shall obtain release approval from Director of Army Technical Information, Office of the Chief of Research and Development, Army Research Office, Washington, D. C.)

4/65–132 **An Automated Sea Surveillance System.** January 1, 1964. N. S. Prywes, W. I. Landauer, L. J. Cohen and D. Lefkovitz. Computer Command and Control Co., Philadelphia, Pa. Rept. no. 4 101 4. Contract Nonr406800. (AD–428 895). 70 pp.; OTS price $7.60.

The organization and maintenance are presented of an automated sea surveillance file for merchant ships. The solution of this problem must be based on a description and definition of the file itself (formats of records in the files and incoming reports) and flow-charts describing the updating processes. However, the problem of sea surveillance is so extensive that to consider all the processes involved is beyond the scope of this investigation. Therefore only the more fundamental processes were considered. The study was conducted as a test case for evaluation of the usefulness of new concepts in computer memory organization generally, and the application of associative memories in particular. (author)

4/65–133 **The Medium-Sized Information Service: Its Automation for Retrieval.** December 30, 1963. Berthold Altmann. Harry Diamond Lab., Washington, D. C. HDL Proj. 01200. HDL TR 1192 (AD–429 242). 26 pp.; OTS price $.75.

A theoretical discussion of the basic elements influencing the organization and operation of a documentation office (the type of information, the clientele, the personnel, and the tools and machines) is followed by a case study. The operations of a medium-size reference service were automated first by utilizing EAM equipment; later a computer. In the process, the coordinate indexing system was replaced by the ABC (Approach-By-Concept) system, permitting manual as well as automatic retrieval. Possible application of this system to large documentation centers is explained. (author)

4/65–134 **On a Study of Information Storage and Retrieval.** May 15, 1964. Interim technical rept. D. S. Sharp and J. E. McNulty. Moore School of Electrical Engineering, Univ. of Pennsylvania, Philadelphia, Pa. MSEE 65 01. Contract Nonr551 40. (AD–602 222). 22 pp.; OTS price $.75.

A preliminary system study was carried out at ASO (Aviation Supply Office) to determine the information needs of personnel and the actions requiring the use of the file for input/output. A study of equipment characteristics is in process and various equipment has been analyzed for suitability. Experimental equipment configurations are being considered as well as commercially available machines. A further study of the detailed information needs in a real-time environment is planned in the Catalog, Stock Control and Technical sections of ASO. (author)

4/65–135 **A Proposal for the Indirect Retrieval of Unpublished Technical Material.** August 1962. R. L. Patrick. RAND Corp., Santa Monica, Calif. P–2616. (AD–605 773). 6 pp.; OTS price: HC $1.00/MF $.50.

The Armed Service Technical Information Agency provides an information retrieval service to Department of Defense contractors with a demonstrated need to know. This paper proposes a subsequent search of the Field of Interest Register if a regular ASTIA bibliographic search fails to produce a stipulated minimum number of items. (author)

4/65–136 **The Use of Electronic Data Processing in Corrections and Law Enforcement.** December 18, 1963. Paul Emil Resta and Charles Philip Smith. System Development Corp., Santa Monica, Calif. Rept. no. SP1495 (AD–427 165). 12 pp.

Presented at American Society of Criminology, Dec. 30, 1963, Cleveland, Ohio. Reviews the reasons, methods, accomplishments and goals of the use of electronic data processing in the fields of correction and law enforcement. Suggests appropriate application and discusses problem areas. Concludes that electronic data processing will permit more effective use of the research, statistical and case history data in building a sounder theoretical base in the field of law enforcement. (author)

4/65–137 **An Information Center for Law Enforcement.** January 2, 1964. Carlos A. Cuadra, Herbert H. Isaacs, Frances Neeland and Everett M. Wallace. Systems Development Corp., Santa Monica, Calif. Rept. no. TM1670. (AD–427 678). 16 pp.; OTS price $1.60.

This paper was prepared at the request of the Subcommittee on Research of the Peace Officers' Association of the State of California. It describes the characteristics of a typical Information Center to serve the members of the Association. The potential services and costs of such a center are discussed. Several recommendations are made for an initial study to define system requirements and, if appropriate, for the tasks necessary to establish the center. (author)

4/65–138 **Design and Implementation of the American Society for Metals Mark II Documentation System.** Clayton A. Shepherd. *Parameters of Information Science.* pp. 347–354. (see 4/65–1)

The ASM Mark II Documentation System is currently being designed and developed. The need for more economical operation, compatibility with other major systems, and the utilization of a controlled natural language thesaurus, are significant factors leading to the redesign of the original ASM-WRU System. (author)

4/65–139 **Quality Identification in a Technical Information System.** Jocelyn Brewer. *Parameters of Information Science.* pp. 247–254. (see 4/65–1).

This paper discusses quality identification for documents entering an information system, presents the Quality Profile as an analysis tool, outlines implied directives to the system, and proposes modes of system implementation. (author)

4/65–140 **A Pragmatic Approach to Retrospective Searching.** Robert E. Swid. *Parameters of Information Science.* pp. 319–322. (see 4/65–1)

It is argued that centralized information centers tend to place a surfeit of emphasis on mechanized file searching. Such capability should be derived as an intrinsic by-product of the primary system complex, and should be gained at the lowest possible cost. Despite the economic and systematic constraints, provision must still be made for the traditional needs of retrieval systems, i.e., optimization of relevancy to a search request, powerful input/output options, continual system evaluation, and an integrated man-machine relationship. (author)

4/65–141 **Investigation of Systems for the Intellectual Organization of Information.** Susan Artandi. *Parameters of Information Science.* pp. 45–47. (see 4/65–1)

Systems included in the study are to be described in terms of their distinctive characteristics. It is assumed that a set of descriptive criteria applicable to information organization systems in general can be created and that sufficiently uniform descriptions can serve as a basis for comparative studies. (author)

4/65–142 **An Executive System Implemented as a Finite-State Automaton.** Roy E. Heistand. CACM, vol. 7, no. 11. November 1964, 669 pp.

The 473L command and control system used by the Air Force permits many operators to access large data files through the use of a computer. The man-machine interface is satisfied by several communication consoles from which operators may enter queries and view replies. A data link permits remote stations to send messages, status reports and inventories directly to the computer. The information received over the online data link is used to update the data files which are stored on disk.

The 473L programming system is divided into an Executive Control Program and five components with different processing priorities. These priorities permit the system to be most sensitive to the console inputs and permit the operators at all the consoles to time share the central processor. The Executive Control Program provides for the orderly transitions of control among the programming system components. The major emphasis of the paper is on the technique of using the definition of a finite-state automaton for organizing the Executive Control Program. (author)

Application of Computers to Classification, Indexing, and Text Processing

4/65–143 **A Programming System for Automatic Classification with Applications in Linguistic and Information Retrieval Research.** October 1964. A. G. Dale, N. Dale, and E. D. Pendergraft. Report LRC–64–WTM–4 of the Linguistics Research Center. University of Texas, Austin, Texas. NSF Grant GN–208. 19 pp.

Two key problems, both of which must be solved as a part of continuing progress toward full use of the computer, concerned the authors: automatic classification techniques, and how to use them as a basis for self-organizing descriptive systems in syntactic and semantic research. Three objectives were discussed: 1. an appropriate basis for retrieval operations in the machine environment; 2. techniques that are computationally feasible on a significantly large scale; 3. immediate use in order to facilitate current research and experimentation with man-machine systems.

Two applications involving the projected use of the programming system are described: a feedback retrieval model in which clumped associations among index words are used for sample output from a document file and user feedback supplies additional data for a reordering algorithm; a proposed procedure for automatic generation of syntactic and semantic descriptions, given primitive symbols of an input text and certain binary relations between symbols as initial data. (authors)

4/65–144 **Is Automatic Classification a Reasonable Application of Statistical Analysis of Text?** August 31, 1964. Lauren B. Doyle. System Development Corporation, Santa Monica, California. Report SP–1753. 34 pp.

The statistical approach to the analysis of document collections and retrieval therefrom has proceeded along two main lines—associative machine searching and automatic classification. The former has been favored because of the tendency in the computer field to strive for new methods of dealing with the literature. Automatic classification study also has been thriving, and some of the reasons for this are discussed.

The crucial question of the quality of automatic classification is treated at considerable length, and empirical data are introduced to support the hypothesis that classification quality improves as more information about each document is used for input to the classification program. Six non-judgmental criteria are used in testing the hypothesis for 100 keyword lists (each list representing a document) for a series of computer runs in which the number of words per document is increased progressively from 12 to 36. Four of the six criteria indicate the hypothesis holds, and two point to no effect. Previous work of this kind has been confined to the range of one through eight words per document. (author)

4/65–145 **Structure Matching in Information Processing.** April 1964. Thesis, Edward Henry Sussenguth, Jr., Harvard University. Published as Report ISR–6, Computation Laboratory of Harvard. (see 1/65–87)

A new technique for comparing structured data is presented. In particular, an algorithm is developed which makes possible the efficient comparison of any two structures which can be represented by a graph; i.e., by a collection of nodes with interconnections

between certain pairs of nodes. Examples of structure-matching in chemical information retrieval, the manipulation of abstract mathematical groups, and linguistic analysis are presented to illustrate the usefulness of this technique in diverse areas of information processing. The proposed matching technique is novel as it avoids the backtracking of the cut-and-try methods currently used. The algorithm (which has been programmed for the IBM 7090) is based upon simple concepts of set and graph theory. Briefly, the procedure consists of determining certain simple properties of the nodes of the structures being tested. Those subsets of nodes of the two structures which exhibit equivalent properties are equated. A standard procedure is then employed to reduce the subsets to other subsets with fewer members, and other procedures are used to generate new subsets from those previously formed. Eventually either each subset is reduced to a single member, thereby delineating an explicit correspondence between the nodes of the structures, or an incompatibility arises, indicating that no correspondence between the structures is possible. (author)

4/65–146 **Common Vocabulary Approaches for Government Scientific and Technical Information Systems.** December 1963. Datatrol Corp., Silver Spring, Md. Technical rept. no. IR10. Contract NSF C342. (D–430 000). 108 pp.; OTS price $9.60.

The present study, which takes a broader look into the vocabulary problem, was given added impetus by the more recent interest of the Committee on Scientific Information (COSI) of the Federal Council for Science and Technology. Specifically, the objective was to develop and document guidelines and criteria for a decision on a single Government-wide approach to an information retrieval vocabulary. The main body of the report centers on a discussion of the various alternatives. In studying the many ramifications of the problem, however, certain ancilliary data were acquired; or developed which appeared to be pertinent to the broader problem of the over-all flow of scientific information; these are presented in the Appendix to the report. (author)

4/65–147 **Adaptive Techniques as Applied to Textual Data Retrieval.** Final rept. August 1964. Douglas Aircraft Co., Inc., Newport Beach, Calif. Contract AF30 602 3134, Proj. 4594, Task 459404 RADC TDR64 206. (AD–605 260). 306 pp.; OTS prices: HC $7.00/MF $1.50.

Document screening is considered as the formation of a classification vector whose coordinates relate the pertinence of the document to various subject fields. This vector is the resultant of vectors assigned to key words and phrases in the document. Adaptive techniques are described which assign vectors to key words on the basis of their occurrence in a corpus of preclassified documents. The results of experiments, conducted to determine the accuracy with which these adaptive techniques approximate key word relevance vectors assigned by expert judgment, are presented. An overall system design philosophy was developed which includes an automatic input mechanism and a device for monitoring the key word relevance vectors to retard system obsolescence. Methods for the hardware implementation of the system are discussed, and computer simulation results on one type of extremely efficient character recognition logic are described. A statistical model was formulated for the analysis of the document screening device. The model is used to predict system error rates as a function of the document corpora, the obsolescence of the relevance vectors, the accuracy of the input device, and the variation of other system design parameters. (author)

4/65–148 **Description of an Algorithm for Documentary Analysis.** December 1963. Y. Lecerf and A. Leroy. EURATOM, Joint Nuclear Research Center Ispra Establishment (Italy), Scientific Information Processing Center (CETIS) Report no. 6. 14 pp.

A paper presented at the First Congress of the French Computing Association (Association Française de Calcul), Grenoble, September 14–16, 1960. In the future systems of automatic documentation each one of the different stages of the classical documentary work ought to be automatized. The authors of this paper develop a method for the automatization of the stage *analysis of documents*. It is based upon the idea that one rational way to express the contents of scientific texts is in diagram form. This transformation must, of course, be carried out by a machine and that in a sufficiently general way which is adaptable to a great number of languages. In this paper the stress is put upon an analysis of French.

It is first demonstrated how a computer can pass to an intermediate representation, to the *stemma*. The next operation consists in establishing the semantic links between certain elements of a stemma or between elements of several stemmas; completed by a research for the solutions of problems of synonymia and polysemia by means of a *general diagram*, this operation provides the representation in diagram form and this, again, leads to a completion of the general diagram.

4/65–149 **The Baseball Program: An Automatic Question-Answerer. Volume II. Flow Charts.** April 11, 1963. A. K. Wolf, C. S. Chomsky and B. F. Green, Jr. Lincoln Lab., Mass. Inst. of Tech., Lexington, Mass. Rept. no. TR306. Contract AF19 628 500. ESD TDR63 598. (AD–446 579). 178 pp.

This volume contains flow charts of the majority of routines within the program. The flow charts are divided into six sections: Executive Routines, Syntatic Analysis Routines, Content Analysis Routines, Processor Routines, Responder Routine, and Other Routines. (author)

4/65–150 **First Congress on the Information System Sciences. Session 17. The Flexibility of Automated Information Systems.** October 1963. MITRE Corp., Bedford, Mass. Rept. no. SS17 Contract AF33 600 39852, Proj. 704. ESD TDR63 474 17. (AD–422 457). 31 pp.; OTS price $3.60.

The common characteristics of information systems are detailed, with systems defined as having procedural restrictions and orderliness. The discrete parameter types of simple man-made systems are compared with biological systems to show the flexibility of the apparently continuous parameter aspect of the latter. Hierarchic arrangements are explained and their uses in formal systems described. The flexibility as well as the inefficiency and limitations of informal systems are shown, and the possible reduction in flexibility when formal systems take over is discussed. The goals and requirements of systems design activity are covered, taking into account the use of automation and mechanization to attain the goals. Advantages and disadvantages of thinking machines are discussed, and their use for military command and control purposes is examined in relation to operational effectiveness. (author)

4/65–151 **Program Documentation for Mark I Statistical Association Procedures for Message Content Analysis.** December 1963. J. Baker, R. Vicksell and J. Spiegel. MITRE Corp., Bedford, Mass. 48 pp. Rept. no. SR89. Contract AF19 628 2390. Proj. 7020. ESD TDR63 259.

A statistical method for automatic document retrieval and message content analysis is described. The method involves building a matrix for the corpus based on word co-occurrences within sentences; this matrix is then normalized in order to eliminate what are considered to be extraneous factors. The normalized matrix is used by the retrieval algorithm to expand a set of query terms to include terms associated with them, the new set, in turn, being used to select documents from the corpus. All of these operations are performed on an IBM 7090 computer. This report give a detailed description of the computer programs involved. (author)

4/65–152 **Operation Manual for the IBM 7090 Exclusive Stratification Program.** September 1963. David Lefkovitz. Moore School of Electrical Engineering, Univ. of Pennsylvania, Philadelphia, Pa. Rept. no. 64 03. Contact Nonr55140. (AD–424–727). 12 pp.; OTS price $3.60.

4/65–153 **Automatic Stratification of Descriptors.** September 15, 1963. 1 vol. David Lefkovitz. Moore School of Electrical Engineering, Univ. of Pennsylvania, Philadelphia, Pa. Rept. no. 64 03. Contract Nonr551 40. (see 7/64–163) (AD–423 647). OTS price $12.00.

A system is suggested that would make use of a digital computer in order to constantly revise the classification of the

descriptors in terms of their actual usage within the document descriptions themselves. A two stage process is intended to produce a two level tree or reference catalog which in the first level presents aggregates of highly associated descriptors, while the second level presents respective aggregates of completely dissociated descriptors. The former is called an inclusive partitioning and the latter, an exclusive partitioning. The paper is principally concerned with the exclusive partitioning, but the last chapter is devoted to a discussion of the complete two stage process where a small selection of descriptors taken from the file of the Armed Services Technical Information Agency (ASTIA) is used to illustrate the construction of a sample index of Thesaurus. (author)

4/65–154 **Automatic Classification for the DDC Mathematics Collection; a Statistical Optimization of Search Time in an Information Retrieval System; the Use of Real Time Computers for Inventory Control.** Research progress rept., May 1964. P. L. Leifer, B. Zimmerman, and D. D. Sharp, Jr. Moore School of Electrical Engineering, Univ. of Pennsylvania, Pa. MSEE–64 17, MSEE–64 21. Contract Nonr551 40. (AD–602 226). 3 pp.; OTS price $.50.

4/65–155 **A Parsing Program for Categorical Grammars.** September 1964. Martin Kay. RAND Corp., Santa Monica, Calif. RM–4283–PR. Contract AF49 638 700. (AD–605 822). 30 pp.; OTS prices: HC $2.00/MF $.50.

The paper describes a computer program for parsing sentences using a categorical phrase-structure grammar. The distinguishing feature of a categorical grammar is that the tags associated with individual words contain all the information necessary to recognize sentences and determine their structures. The familiar table of grammatical rules has no analog within this system. The grammatical tags of a categorical grammar must have a definite structure of their own. The greater part of the paper is given to a discussion of this structure and an algorithm that exploits it. Various ways of representing tags on paper and in the computer are considered, together with corresponding variants of the algorithm.

4/65–156 **Qualitative and Quantitative Procedures in Content Analysis.** December 15, 1964. Alexander L. George. RAND Corp., Santa Monica, Calif. 38 pp. P–617. (AD–604 628). 38 pp.; OTS prices: HC $2.00/MF $.50.

In the specialized terminology of propaganda analysis, it is useful to distinguish between statements confined to content description and those which constitute inferences from content about its antecedent conditions—e.g., propaganda goals, elite policy calculations, actual events. To avoid the confusion surrounding the distinction between quantitative and qualitative content analysis in the past, a new distinction is proposed, i.e., that between "frequency" and "non-frequency" content analysis: (a) If the propaganda analyst believes that the number of times a given content characteristic appears is significant for purposes of inference, this characteristic is regarded as a frequency indicator; (b) however, if he decides that the mere presence or absence of a given content characteristic is significant, then it is regarded as a non-frequency indicator.

4/65–157 **Mechanized Documentation: The Logic Behind a Probabilistic Interpretation.** April 1964. M. E. Maron. RAND Corp., Santa Monica, Calif. Rept. no. P2898. (AD–437 781). 20 pp.

4/65–158 **CLASSIFIER. An Automated Computer-Oriented Information Classification System.** February 1964. John B. McLean and Edward Morenoff. Rome Air Development Center, Griffiss Air Force Base, N.Y. RADC TDR64 14. (AD–434 768). 17 pp.; OTS price $1.60.

CLASSIFIER is an automated computer-oriented information classification system which provides a library structure for information as well as document retrieval. The information content of the document is classified at levels of detail ranging from the document to a sentence. Statistically determined measures of relevance indicative of the degree of association of each unit of information with respect to each class of the hierarchy are defined. The more frequently a word is used in a class, rather than in general usage, the greater the association with that class. Conversely, the less frequently a word is used in a class, rather than in general usage, the less association with that class. The result is a series of lists of units of information for each class of the hierarchy, rated in order of decreasing relevance. Inherent in the classification process is the ability to select index terms for the units of information to complement the hierarchical structure. Nonstatistical measures can supplement the statistical techniques if desirable. (author)

4/65–159 **Answering English Questions by Computer: A Survey.** April 2, 1964. R. F. Simmons. Systems Development Corp., Santa Monica, Calif. Rept. no. SP1556. Contract SD97. (AD–437 610). 64 pp.; OTS price $6.60.

Fourteen question-answering systems which are more or less completely programmed and operating are described and reviewed. The systems range from a conversation machine to programs which make sentences about pictures and systems which translate from English into logical calculi. Systems are classified as data based, text based, and inferential. Principals and methods of operations are detailed and discussed. It is concluded that the data base question answerer has passed from initial research into the developmental phase. The most difficult and important research questions for the advancement of general purpose language processors are seen to be concerned with measuring meaning, dealing with ambiguities, translating into formal languages and searching large tree structures. (author)

4/65–160 **Research in Document Classification and File Organization.** November 13, 1963. Harold Borko. Rept. no. SP 1423. (AD–425 531). 12 pp.; OTS price $1.60.

In an information storage and retrieval system, classification provides a means for organizing a mass of material into groups so that related items are brought together in a systematic fashion. By grouping documents into categories, the number of items to be scanned in response to a search request can be reduced and the efficiency of the system increased. To design a classification system, one must specify the number of classes to be established and the principle to be used in determining class membership. A number of mathematical procedures have been suggested for devising classification schedules. These include factor analysis, clump theory, and latent class analysis. It also has been suggested that these or similar techniques can be used to automatically classify documents into correct categories and speed up the processing of incoming material. The initial results appear promising. Further research is being undertaken to determine the retrieval effectiveness of automated indexing and classification systems. (author)

4/65–161 **Computer Analysis of Text for the Measurement of Social Perception During Human Relations Training.** January 9, 1964. J. D. Ford, Jr. System Development Corp., Santa Monica, Calif. Rept. no. SP1373 001 00. (AD–430 260). 11 pp.; OTS price $1.60.

This report describes a preliminary study where automated content analysis was applied to the assessment of changes in social perception during sensitivity training. It describes the particular type of training studied, and its procedures and goals; the computer techniques by which automated content analysis was accomplished; the type of data used to infer social perception; and the results of the content analysis. (author)

4/65–162 **Automatic Paraphrasing in Essay Format.** July 21, 1964. Sheldon Klein. System Development Corp., Santa Monica, Calif. Rept. no. SP1602 001 00. (AD–443 744). 60 pp.

This report describes an operating computer program that accepts as input an essay of up to 300 words in length, and yields as output an essay-type paraphrase that is a nonredundant summary of the content of the source text. Although no transformations are used, the content of several sentences in the input text may be combined into a sentence in the output. The format of the output essay may be varied by adjustment of program parameters, and the system occasionally inserts subject or object pronouns in its paraphrases to avoid repetitious style. (author)

4/65-163 **Retrieval of SIE Reports Using the Protosynthex System.** December 9, 1963. Phyllis R. Kennedy. Rept. no. TM1635. (AD-427 823). 61 pp.

Preliminary results are presented of an information retrieval experiment using SDC's Protosynthex System and an experimental corpus provided by the Science Information Exchange. Retrieval runs were carried out using a number of program options. The options which appear to provide the most effective retrieval of SIE reports are indicated. Analysis of the experimental results is proceeding. (author)

4/65-164 **Synthex 1964: A Progress Report and a Research Plan.** May 6, 1964. Robert F. Simmons. System Development Corp., Santa Monica, Calif. Rept. no. TM1807. Contract SD97. (AD-443 756). 33 pp.

This report summarizes accomplishments of the Synthex natural language processing project and presents a research plan for the immediate future. It references and annotates Synthex documents. Past work has been concerned primarily with indexing and retrieval logic, the identification of word-classes without large dictionaries, and the development of several forms of syntactic analysis and synthesis programs. Current and future work is oriented: 1. toward the development of a context sensitive parser which can give relatively unambiguous analyses of sentences; 2. the development of decision procedures for discovering semantic codes and the discovery of rules of combination for semantic codes; and 3. the continued development and improvement of Protosynthex as a question-answering system whose capabilities will expand to include features of logical and linguistic interference. (author)

4/65-165 **Controlo f Style with a Generative Grammar.** July 24, 1964. Sheldon Klein. System Development Corp., Santa Monica, Calif. Rept. no. SP1633 001 00. (AD-443 742). 24 pp.

Certain parameters of style have been incorporated in an immediate constituency general grammar which is itself a component in an automatic paraphrasing system. The parameters actually tested in generation experiments are those of syntactic structure and frequency of occurrence of pronouns. The program can be modified to control additional quantitative measures of style. The complete paraphrasing system performs both an immediate constituency and a dependency analysis of an input text and uses these to monitor the output of a generation grammar as part of the task of producing an essay-type paraphrase of the source text. (author)

4/65-166 **Natural Language Computer Processing of Los Angeles Police Department Crime Information.** Progress rept. no. 1. April 1, 1964. Herbert H. Isaacs, William O. Crossley and Charles R. Withers. System Development Corp., Santa Monica, Calif. Rept. no. TM1793 000 00. (AD-440 071). 65 pp.

This is the first in a series of reports decribing a year-long program of investigation and experimentation in computer processing of crime information in natural Engilsh language. It provides some of the background information concerning the research effort. and relates the progress to date. It is organized as follows: A brief summary; the need for this research and the previous work that led to its initiation; outlines of the entire project; summary of the activities completed thus far; a description of specific tasks within the overall effort in terms of their purpose, activity schedule, and work completed. (author)

4/65-167 **A Preface to Computational Stylistics.** February 17, 1964. Sally Yeates Sedelow and Walter A. Sedelow, Jr. System Development Corp., Santa Monica, Calif. Rept. no. SP1534. (AD-433 513). 24 pp.; OTS price $2.60.

Stylistic analysis, the study of patterns formed in the process of the linguistic encoding of information, is of importance to any major research focused upon or dependent upon the production or analysis of language. Through the use of computers, it should be possible to achieve more accurate detection and delineation of such linguistic patterns than has hitherto been the case. Computational stylistics has immediate, practical implications for work in areas ranging from machine translation and automatic abstracting to the social sciences and humanities. For adequate machine translations and automatic abstracts, algorithms of normative style for the textual genre being translated or abstracted must be available; the use of the computer for stylistic analysis will help make possible the recognition and specification of such algorithms. Stylistic analysis is also integral to the detection of idiosyncratic uses of language which distinguish one author from another. (author)

4/65-168 **Some Compromises between Word Grouping and Document Grouping.** Lauren B. Doyle. System Development Corp., Santa Monica, Calif. March 26, 1964. Rept. no. SP1481. (AD-440 044). 22 pp.

Statistical analysis of the text of document collections has yielded, for information retrieval purposes, two broad classes of output: word grouping and document grouping. Associative indexing comes under the general heading of word grouping and automatic classification is a kind of document grouping. Document grouping and word grouping can be combined to give a scheme of classification with more attractive features than could be achieved with either document grouping or word grouping alone. A hierarchical grouping program written by J. H. Ward of Lackland Air Force Base for use in classifying personnel by skill and aptitude is discussed. This program turns out to be nearly ideal as a basis for a mixed document and word-grouping approach. (author)

4/65-169 **Aerospace Intelligence Data System (AIDS).** Quarterly rept. no. 8, May 16-August 15, 1962. Thomas J. Watson, Research Center, Yorktown Heights, N. Y. Contract AF 19 626-10. (AD-423 305). 219 pp.; OTS price $5.60.

Progress on the AIDS Research Contract is presented. The research results of ten tasks covering nonnumerical information processing problems for the period May 16, 1963 to August 15, 1963 are reviewed and presented in the main text. Two appendices describe experiments with citation indices and applications of statistical programs. (author)

4/65-170 **Automatic Abstracting of Russian Text.** Final rept. June 1964. TRW Computers Co., Canoga Park, Calif. TRW 432 402. Contract AF30 602 3202, Proj. 4594, Task 459401. RADC TDR 64 155. (AD-602 357). 104 pp.; OTS prices: HC $4.00/MF $.75.

4/65-171 **Concept Analysis by Syntax Processing.** Robert S. Ledley and James B. Wilson. *Parameters of Information Science.* pp. 1-8. (see 4/65-1)

Mechanical language translation systems and heuristic programming are applied to the problem of concept recognition on both the deductive and inductive approaches. The paper presents a new application and integration of these modern programming techniques. (authors)

4/65-172 **Accumulation of Natural Language Text for Computer Manipulation.** Lee Ohringer. *Parameters of Information Science.* pp. 311-313. (see 4/65-1)

The University of Pittsburgh Computation and Data Processing Center is developing a system to record currently published technical literature in computer-compatible form. Using the paper tapes which are punched during the printing process for Linotype, Monotype, and photo-typesetting, this system will record total information, including text, type font, type size, and page arrangement. Presently two IBM 1401's, a 7070 and a 7090, and a PDP 4 are available for this project. (author)

4/65-173 **Machine-Set Chemical Structures.** Joseph H. Kuney and Bernard G. Larorchak. *Parameters of Information Science.* pp. 303-305. (see 4/65-1)

Photocomposition of chemical structures has been achieved. The method seems suited to computer processing and this phase of the research is now under study. The procedure followed will be described and samples shown. (authors)

4/65-174 **High-Speed Text Composition.** John W. Seybold. PRMG, November 1963. pp. 64-65.

This is a lucid and realistic survey of the current state of the art, and the immediate future, of computer applications to

the rapidly changing, ancient art of printing. Such varied developments as character-recognition, phototypesetting and teletypewriter transmission are discussed in an evaluation of the prospects promised by their intelligent combination (with computers and automated processes) to create a revolution in the many-linked chain from typescript to printed sheet. (CPGR)

4/65–175 **Automatic Book Indexing by Computer.** Susan Artandi. AMDO, vol. 15, no. 4. October 1964, pp. 250–257.

This paper describes the development and evaluation of an automatic method of deriving suitable index entries from natural language text. A section of a chemistry textbook was selected as the experimental document. Both qualitative and quantitative measures were developed to evaluate quality of indexing. Evaluation of the experimental index was based on these measures and on a comparison with the average published manually produced index found in the same type of material. (author)

4/65–176 **Computer Interpretation of English Text and Picture Patterns.** Russell A. Kirsch. IREL. Vol. EC-13, No. 4. August 1964. pp. 363–376.

This paper considers a class of information sources consisting of text and pictures. The text is English language text appearing in scientific and technical documents. The picture sources are the largely schematic pictures that occur in the same class of documents. For a tiny fragment of English, the paper shows how the syntactic structure of text may be described, and then goes on to suggest that a similar analysis may be performed on the class of pictures under study. The description of these two kinds of information sources with a single class of descriptive techniques is suggested as an alternative to the synthetic approach in which artificial languages are specified and then learned and used. The major reason for doing syntactical analysis of such sources discussed here is that several information processing operations, amounting to the interpretation of the information sources looked upon as languages, can be done by the technique of syntax direction which uses the results of syntactic analysis to mediate subsequent processes for manipulating the information tokens. The paper concludes with an illustration of an algorithm for matching the sentences given by a simple grammar against the class of simple pictures which these sentences purport to describe. (author)

4/65–177 **Some Flexible Information Retrieval Systems Using Structure Matching Procedures.** G. Salton and E. H. Sussenguth, Jr., *AFIPS Conference Proceedings, Spring Joint Computer Conference*, 1964. vol. 25, pp. 587–597. (see 1/65–87)

The comparison between stored information identifications and requests for information is one of the principal tasks to be performed in automatic information retrieval. In many retrieval systems it has been found necessary or expedient to use complicated constructs for the identification of information. A complete identification for a document or an item of information is often represented by a graph, consisting of nodes and branches between nodes, to identify respectively the key words and relations between key words. The matching of such information graphs with graphs representing requests for information is a relatively complicated operation, particularly since the request structure can be made to match the information structure only partially and incompletely. An efficient topological structure-matching procedure, which does not depend on a specific ordering of the nodes, nor on the presence or absence of certain specified substructures, is described. The procedure is applied to the matching of document graphs with request graphs, and to a retrieval system based on the comparison of syntactically analyzed document excerpts with a stored phrase dictionary. (IBMJ)

4/65–178 **Mechanized Indexing Methods and Their Testing.** John O'Connor. JACM, vol. 11, no. 4. October 1964. pp. 437–449.

Methods of mechanized indexing (subject indexing by computer) which have been proposed are systematically summarized. Every suggested method consists of some document preparation process (mostly or wholly mechanical) followed by the application of indexing rules to the prepared document. A comprehensive document preparation is described from which proposed methods can be derived by selection. It includes full text input, document place (title, abstract, etc.) marking, sentence and paragraph marking, pronoun replacement, and other syntactic marking. It also includes addition of thesaurus headings, position numbers, weighted frequencies, closely associated expressions, importance measures, and reference information. Three kinds of indexing rules are then distinguished and illustrated.

Several general comments on mechanized indexing include remarks on the argument that good mechanized indexing is not feasible, and the argument that mechanized indexing has the advantage, compared to human indexing, of consistency. Some problems of testing mechanized (or any other) indexing quality by the quality of the retrieval it permits are described. Index duplication studies are suggested as an alternative kind of empirical investigation of mechanized indexing methods. (author)

4/65–179 **Some Patterns Observed in the Contextual Specialization of Word Senses.** John C. Olney. ISSR, vol. 2, no. 2. July 1964. pp. 79–101. (see 7/64–159)

4/65–180 **A Concordance Generator.** K. F. Scharfenberg, P. H. Smith, Jr., and R. D. Villani, IBMJ, vol 3, no. 1. 1964. pp. 104–111.

The structural design of a general purpose program for concordance preparation is described. Options in the input format, the operating mode, and the output edit provide wide flexibility in organizing data in a form convenient for many analytical purposes. An experimental program was written, and some results obtained in testing the program are included. (IBMJ)

4/65–181 **Typesetting in the Computer Age.** August 1964. *Print in Britain.* 12(4) Supplement pp. 1–8.

A report of the Computer Typesetting Conference organized by the Institute of Printing at London University in July 1964.

4/65–182 **New Computers Are Changing Printing Industry Production Techniques.** John N. Pannullo. IPAL. June 1964. pp. 60–61.

4/65–183 **Computers Could Lead to New Standards of Typesetting.** Alexander Lawson. IPAL. June 1964. pp. 70–71.

4/65–184 **A Document Retrieval System for Man-Machine Interaction.** G. Salton. The Computation Laboratory of Harvard University, Cambridge 38, Massachusetts.

An automatic document retrieval system, programmed for the IBM 7094, is described. The system is designed to process English texts and search requests, and uses statistical, syntactic and semantic procedures for the analysis of information and the identification of relevant items.

The operations are planned around a central supervisor, which in turn calls on the various subroutines, as desired. This organization makes it possible to alter both the processing sequences and the matching criteria between stored information and search requests, thus producing a variable amount of information in response to a given search request. The system therefore lends itself to interaction with the user by enabling the latter continuously to change his search requirements. Furthermore, the various analytical procedures may be evaluated by comparing the retrieval results obtained under different processing conditions. (author)

Subject and Linguistic Analysis – General

4/65–185 **Study of Theories and Models of Information Storage and Retrieval. Report No. 7: Graphs and Algorithms for Term Relations.** July 30, 1964. Donald J. Hillman. Center for the Information Sciences, Lehigh University, Bethlehem, Pa. NSF grant GN–283. 21 pp.

An algorithm is presented for grouping the terms of a document corpus into genera, and it is shown that the different genera of the graph G = (T, M), where T is the set of index terms and M the reflexive, symmetrical relation of matching, constitute a partition of T. An effective method is next described for discovering the orderings of terms within each genus. It is then shown

that every genus is a topological space, and a technique is formulated for assigning neighborhoods to terms. An algorithm is next described for forming the concepts contained in a given genus, so that the formal theory of retrieval described in earlier reports is now susceptible of empirical corroboration.

4/65–186 **An Evaluation of Links and Roles Used in Information Retrieval.** Master's thesis. July 1964. Jefferson D. Sinnett. Air Force Materials Lab., Research and Technology Div., Wright-Patterson AFB, Ohio. Proj. 7381, Task 738103. ML TDR64 152. (AD–606 192). 154 pp.; OTS prices: HC $5.00/MF $1.00.

Links and roles are techniques for modifying a basic coordinate indexing method. The purpose of the modification is to reduce the amount of irrelevant information retrieved without losing relevant information. Theory of retrieval is reviewed, mathematical models are discussed, retrieval is described in terms of basic concepts for algebra of sets, and a method of scoring effectiveness is presented. Roles were found inferior to links. Links showed reduction of irrelevant information by over 56% while incurring less than 5% loss of relevant information. (author)

4/65–187 **Investigation of Search Metric Capabilities.** September 1963. J. F. Rial. Mitre Corp., Bedford, Mass. Rept. no. W5340. Contract AF33 600 39852, Proj. 438L. ESD TDR63 167. (AD–420 266). 11 pp.

Earlier work at MITRE defined a distance function for document retrieval systems queried by Boolean combinations of Keyworks. A plan for systematically exploring the retrieval capabilities of that function is described. (author)

4/65–188 **Fact Correlation Experimentation.** May 1964. N. H. Hardwick, S. N. Jacobson, D. E. Rogers, and G. H. Woolley. Radio Corp. of America, Bethesda, Md. Contract AF30 602 2979, Proj. 4594, Task 459401. RADC TD64 190. (AD–603 697). 181 pp.; OTS prices: HC $5.00/MF $1.00.

The report covers several facets of an experimental program to develop and test techniques to the realization of fact correlation capability. Primarily, these techniques are linguistically oriented and particularly are concerned with paragraph analysis, the ability to reorganize natural language text in a meaningful way using automatic and semiautomatic methods. Nonlinguistic approaches based on vocabulary manipulation were also investigated to a limited extent. The test vehicle measures their performance in reassembling previously segmented documents. To provide data for the specification of the techniques a sample of scientific documents was prepared for machine processing. A similar sample containing documents matching those in the analysis sample is available for testing the ability of the techniques to reassemble documents. Computer programs representing a document reassembly system to test the linguistic techniques were written in the COMIT programming language for the IBM 7090 computer. Programs for document reassembly to test other techniques were implemented on the RCA 301 computer. Each of these systems is described in detail. (author)

4/65–189 **Keyword-in-Context (KWIC) Indexing: Background, Statistical Evaluation, Pros and Cons, Applications.** John Michael Sedano. Master's thesis. Pittsburgh Univ., Pa. 1964. (AD–443 912). 77 pp.

This thesis discusses a new method of machine indexing literature. The author assumes that the reader is unfamiliar with the new discipline of Information Science. The background of machine indexing is developed deductively. Next, the author describes KWIC indexing, evaluates statistically the hypothesis upon which it is based (that titles are descriptive of the article or document they represent), lists its pros, cons and applications. Finally, the author concludes that KWIC indexing is an efficacious new tool. (author)

4/65–190 **The Necessity of Non-Descript Descriptors.** R. A. Fairthorne. *Parameters of Information Science*. pp. 355–356. (see 4/65–1)

The notational properties that discriminate and identify items within retrieval systems must satisfy various combinatorial requirements for efficiency. These conflict with the requirements for linguistic expression of notional requirements by the human invoker. Rarely can notation directly reflect notion. The paper illustrates, develops, and applies this theme. (author)

4/65–191 **Descriptors, Terms, Subject Headings, Term-Indexing, Subject-Indexing—Definitions.** E. H. Brenner. *Parameters of Information Science*. pp. 387–388. (see 4/65–1)

Descriptors are individually undefinable as *terms* or *subject headings* without considering their treatment. If a system consistently calls for term-indexing and the vocabulary has been so formed, the descriptors may be called *terms*; if the system consistently calls for subject-indexing, the descriptors may be called *subject headings*. The American Petroleum Institute system which allows for both term-indexing and subject-indexing of the same material is briefly described. (author)

4/65–192 **Role Indicators and Their Use in Information-Searching-Relationship of ASM & EJC Systems.** Marjorie R. Hyslop. *Parameters of Information Science*. pp. 99–107. (see 4/65–1)

A comparison is made of the meanings of role indicators in the ASM and EJC systems. Studies of *popularity* of role indicators were conducted on 150 operational ASM searches. The popular roles are analyzed in relationship to the EJC roles. Results shed light on the compatibility of the two systems and the general efficacy of this indexing device. (author)

4/65–193 **Classification Research Group, Bulletin No. 8.** *J. Doc.*, vol. 20, no. 3. September 1964. pp. 146–169.

Summarizes discussions and minutes for 1961 to 1963 under the following main foci of interest: development of the Cranfield index-testing project; content analysis, indexing, permutation of facets, and chain indexing; integrative levels as a basis for a general classification; and problems of drafting a general classification, suported by a NATO grant. Appended is a bibliography of writings by members of the Classification Research Group. (D.C.W.)

4/65–194 **Measuring the Reliability of Subject Classification by Men and Machines.** Harold Borko. AMDO vol. 15, no. 4. October 1964. pp. 268–273.

Procedures for measuring the consistency of document classification are described. Three subject specialists classified 997 abstracts of psychological reports into one of eleven categories. These abstracts were also mechanically classified by a computer program using a factor-score computational procedure. Each abstract was scored for all categories and assigned to the one with the highest score. The three manual classifications were compared with each other and with the mechanical classifications, and a series of contingency coefficients was computed. The average reliability of manual classification procedures was equal to .870. The correlation between automatic and manual classification was .766. (author)

4/65–195 **Subject Heading and Facet Analysis.** S. R. Ranganathan. *J. Doc.*, September 1964, vol. 20, no. 3. pp. 109–119.

After establishing the terminology, shows how the choice of the name of the subject of a document and the rendering of the name in the heading of the specific subject entry can be done by facet analysis based on postulates and principles. After showing that subject headings constitute an artificial language, points out that using facet analysis for subject heading does not amount to using class number. Marks out the area for an objective statistical survey of sought heading for subject entry. Calls on Council for Library Resources Incorporated to provide for this project. (author)

4/65–196 **Language-Analysis Problems in the Computer Processing of Natural Text.** W. D. Climenson. IEEW, vol. EWS-6, no. 2. December 1963. pp. 72–78. (BATR)

The goals of computer processing of natural-language text include automatic retrieval of data, automatic indexing, automatic formatting of data for subsequent logical and manipulative

processing. A number of possible techniques are suggested to achieve these goals.

4/65–197 **Introduction to Semantics.** Adam Schaff. 1962. Pergamon Press, New York. 395 pp.

Discusses linguistics, logic and philosophical aspect of the communication process, the meaning of words, and the communicative function of language.

Specific Classification and Indexing Systems

4/65–198 **The Universal Decimal Classification.** Jack Mills. Vol. 1 of the *Rutgers Series on Systems for the Intellectual Organization of Information.* Susan Artandi, Ed. 1964. Graduate School of Library Service, Rutgers, New Brunswick, N. J.

4/65–199 **Radiation Effects Information Center Dual Coordinate Index for Accession Lists.** Rept. for July 1, 1962–September 30, 1963. Radiation Effects Information Center, Columbus, Ohio. October 15, 1963. 1 vol. Rept. no. REIC Accession lists 60 thru 68 (AD–420 135). OTS price $7.60.

4/65–200 **Philosophy and Guidelines for Revision of the Thesaurus of ASTIA Descriptors.** November 1, 1961. T. L. Gillum, P. H. Klingbiel, C. N. Mooers, and E. Wall. Defense Documentation Center, Alexandria, Va. rev. October 1964. (AD–606 639). 24 pp. OTS prices: HC $1.00/MF $.50.

In response to many inquiries from users of the Thesaurus of ASTIA Descriptors and from others interested in the use of controlled vocabularies for mechanized information retrieval, this paper is offered as an outline of the general plan to be followed in the preparation of a Second Edition. The philosophies of the descriptor and thesaurus approaches to information retrieval are discussed, with particular emphasis on the relationships among descriptors. Although this document was intended as a guide line for individuals who had been invited to participate in the preparation of the Second Edition, the discussion of the thesaurus philosophy is believed to be of general interest to documentalists. Included is a bibliography which cites papers dealing with the general concept of technical vocabularies. (author)

4/65–201 **Practical Aspects Concerning the Development and Use of ASTIA's Thesaurus in Information Retrieval.** 1963. J. F. Caponio and T. L. Gillum. Defense Documentation Center, Alexandria, Va. (AD–420 503 and AD–420 843). 19 pp.

A detailed description on the nature of ATIA's report literature control problem and the evaluation of a mechanized system for storage and retrieval of document references pertaining to science, technology, and medicine. The application of the thesaurus concept for control of an indexing vocabulary is related and the revision of the First Edition of the Thesaurus of ASTIA Descriptors is explained. Problems arising from efforts to evaluate terminology and indicate relationships among selected descriptors are discussed in terms of the broad principles derived. Emphasis is placed on the importance of the human element in a mechanized information retrieval system. (author)

4/65–202 **Wadex Word and Author Index. A New Tool in Literature Retrieving.** 1964. E. A. Ripperger, H. Wooster and S. Juhasz. Texas University, Austin, Texas. Grant AF AFOSR33 64 AFOSR 64 1033 (AD–440 961). 6 pp. Reprint from *Mechanical Engineering,* March 1964. pp. 45–50. (Copies supplied by DDC.)

4/65–203 **Fort Detrick Thesaurus Project.** General Electric Co., Washington, D. C. Contract DA18 064AMC132A. OTS price $1.10.

4/65–204 **A New Pharmacological Classification and Coding System with Official Recognition.** Gerald S. Savitz. *Parameters of Information Science.* pp. 89–97. (see 4/65–1)

Based on the recommendations and adoption of the category descriptions for "basic drugs" by the United States Pharmacopeia, XVII Revision (1965), a big step has been taken towards the standardization of pharmacologic terminology. From this has developed an expanded pharmacological classification to include experimental drugs and a broad, basic pharmacological coding system for information storage and retrieval purposes. This new system of pharmacologic classification and coding has been used in an index card service—Drugs in Prospect—since January, 1964. This service is the result of the pharmaceutical industry's request for a quick, thorough and systematic procedure to bring to their attention the best of the current literature reporting on the *first* reference to a pharmacologic activity (animal or human) of a new chemical compound. (author)

4/65–205 **GREMAS—A Method of Classification and Documentation for Organic Chemistry.** R. Fugmann, W. Braun, and W. Vaupel. NADO, vol. 14, no. 4, December 1963, pp. 179–190. (In German.) GREMAS—ein Weg zur Klassifikation und Dokumentation in der organischen Chemie.

4/65–206 **An IGY/IGC Data Publication with a Built-In Retrieval System.** Gertrude London. TAGU, vol. 45, no. 3. September 1964. 445 pp.

A classification and coding system is described which was incorporated into the publication of IGY and IGC meteorological data on microcards. It was derived from inherent and recurrent characteristics of the material and of the chosen publishing medium. The data, nearly 8 million world-wide synoptic meteorological observations, were reproduced on microcards grouped in regular series of basic layouts. The main features were expressed in simple numerical codes which enable users to locate data easily and rapidly on any of over 20,000 IGY and IGC microcards. (author)

4/65–207 **Locating Engineering Literature.** R. A. Jacobson. *Machine Design,* vol. 36, June 18, 1964, pp. 152–158.

Describes the new Engineers Joint Council's Thesaurus, coordinate indexing and the EJC system of roles.

4/65–208 **A New Tool in Literature Retrieving.** E. A. Ripperger, H. Wooster, and S. Juhasz. *Mechanical Engineering,* vol. 86, March 1964, pp. 45–50.

WADEX, standing for Word and Authors in*DEX,* is a new type of computer index for retrospective search rather than for current awareness. The titles are printed fully with a descriptor (word or author's name), reference number, and arranged in two columns. The procedure followed in preparation of this index is described. The first issue of this index is for the 1962 volume of *Applied Mechanics Review.*

4/65–209 **Rotadex—A New Index for Generic Searching of Chemical Compounds.** Irving H. Sher, John O'Connor, and Eugene Garfield. JCHD, vol. 4, no. 1, January 1964, pp. 49–53.

Rotadex consists of a rotated index with molecular formulas, generic structural codes, and the addresses where the references were made to the compounds. Searches are best performed with Rotadex by first defining the minimal, elemental requirements and parameters of acceptable structural codes.

4/65–210 **Classification Schemes for the History of Science.** Magda Whitrow. *J. Doc.,* vol. 20, no. 3. September 1964, pp. 120–136.

In preparation for a cumulative bibliography based on the 87 classified Critical Bibliographies published in *Isis,* the History of Science Society has found it necessary to develop a more detailed classification scheme than had been used. George Sarton's scheme for *Isis* showed several developments during the early years of this century, and he used a more detailed systematic scheme from 1926 until 1952 when a committee of the Society adopted simplifications. Six recent systematic arrangements are mentioned as are a few problems in classifying history of science literature. A new draft scheme is proposed by the author for the annual critical bibliography in *Isis* as well as for the over 100,000 cumulated bibliographic notices. (D.C.W.)

Coding and Notation

4/65–211 **A Notation for Representing Conceptual Information. An Application to Semantics and Mechanical English Paraphrasing.** October 29, 1963. Ross Quillian. System Development Corp., Santa Monica, Calif. Rept. no. SP1395. (AD–425 486). 59 pp.; OTS price $7.60.

This report describes a symbolic notation to be used for representing concept-like information. The notation is designed to permit representation, in a single homegeneous code, of an extremely broad range of conceptual information. The aim of the notation is to allow information to be stored and processed in a computer in a manner that is functionally equivalent, for at least some purpose, to the way humans store and process conceptual information. This would facilitate mechanical simulation of those functions people perform on the basis of their conceptual worlds. The concepts which the notation is presently used to represent are word meanings. Word meanings are concepts made up of extremely large and varied bodies of information, but much of the information comprising one such concept recurs in other work meanings. Thus, the notation is organized into a code such that what actually becomes stored in a computer memory is not a representation of each word itself but rather a non-redundant, grammar-like store and a processing routine, which together can generate a great many different pre-stored word meanings. These give a computer the capability of generating all or any part of the conceptual information making up the meaning of a word, as needed. (author)

4/65–212 **A Coding Format and Machine Method for Specific and Generic Coding of Lists of Chemical Names.** January 1963. J. Frome, J. F. Caponio, P. H. Klingbiel and P. T. O'Day. Defense Documentation Center, Alexandria, Va. (AD–420 579). 9 pp.; OTS price $1.10.

The system described here is as used in the Armed Services Technical Information Agency (ASTIA). ASTIA uses a large thesaurus for classifying and searching technical subject matter which is a combination of terms that encompass all of technology. An attempt is currently underway to break out several art-oriented, highly specific microthesauri. This paper describes the procedure used to help reorganize and refine one of these thesauri, the Chemical Thesaurus. (author)

4/65–213 **Development of Chinese Ideographic Composing Machine.** July 1964. Pittsburgh University, Pittsburgh, Pa. Contract AD36 034AMC3785X AMC TIR12 3 1 1 (AD–447 613). 12 pp.

This report details Army development of a photocomposing machine for Chinese ideographic characters for the preparation of printed propaganda, orientation, and other material. Development of such a machine was found to be feasible in 1954 when studies revealed constants in the execution of Chinese ideographic composition. The current prototype has a vocabulary of 2,500 ideographs developed from a keyboard of 21 basic strokes, 20 ideographic entities, and 11 punctuation marks. The machine's capacity can be expanded to 10,000 ideographs. It operates at 600 strokes per minute and is 66 inches long. 29 inches deep, and 60 inches high. It weighs 2,000 pounds, is housed in a standard S–141 equipment shelter, and is transportable by truck or helicopter. (author)

4/65–214 **Error-Protected Names.** December 1963. Walter J. Huebner, Jr. Behavioral Sciences Lab., Aeronautical Systems Div., Wright-Patterson Air Force Base, Ohio. Contract AF33 657 7362, Proj. 7184, Task 718404 AMRL TDR63 83 (AD–430 714). 22 pp.; OTS price $.75.

A method of generating error-protected encoded information or error-protected names was originated. Any single name resulting from the application of this system is unique in that at least two of its symbols differ from the symbols in corresponding positions of every other name. This property permits the detection of all errors involving single substitution or erroneous interchange of two symbols. In addition, this system will detect well over 97% of all other errors that might occur in transcribing a name. Name lists generated by the method described, are particularly applicable to identification numbers, license plates, parts lists, etc., where mistaken identification is likely to be costly. For example, with an alphabet of 23 symbols, it is possible to generate a list of 6,400,000 error-protected, six-symbol names. The method of generating the lists is described in detail and the error-protective properties of such lists demonstrated mathematically. (author)

4/65–215 **Computer Generation of Atom-Bond Connection Tables from Hand-Drawn Chemical Structures.** W. E. Cossum, M. E. Hardenbrook and R. N. Wolfe. *Parameters of Information Science*. pp. 269–275. (see 4/65–1)

This paper describes preliminary results of an experimental system for input of chemical structures to a large chemical information storage and retrieval complex. A general purpose computer recognizes element and bond symbols and builds an atom connection table from a video image of hand-drawn structural formulas entered into the computer via an optical scanner. The system is discussed in terms of its possible support of present and future CAS services. Input would be generated primarily from the formula indexing operation at CAS and the resultant connection table would be the basic tool for registration of new compounds, substructure searching, linear notation generation, and graphic display. The method of generating the connection table from the video pattern is covered and illustrated by samples of hand-drawn structures. Preliminary economic data are included. (authors)

4/65–216 **Nonrandom Binary Superimposed Codes.** W. H. Kautz and R. C. Singleton. IETT, vol. IT–10, no. 4. October 1964. pp. 363–377.

A binary superimposed code consists of a set of code words whose digit-by-digit Boolean sums $(1 + 1 = 1)$ enjoy a prescribed level of distinguishability These codes find their main application in the representation of document attributes within an information retrieval system, but might also be used as a basis for channel assignments to relieve congestion in crowded communications bands. In this paper some basic properties of nonrandom codes of this family are presented, and formulas and bounds relating the principal code parameters are derived. Finally, there are described several such code families based upon 1. q-nary, conventional error-correcting codes; 2. combinatorial arrangements, such as block designs and Latin squares; 3. a graphical construction; and 4. the parity-check matrices of standard binary error-correcting codes. (authors)

4/65–217 **The Location of Figures in Alpha-Numeric Codes.** R. Conrad. ERGO, vol. 5, no. 2. April 1962. pp. 403–406.

In designing a six-character alpha-numeric code, a combination of operational and scientific factors required two characters to be figures and four to be letters. It was predicted that when the figures were located in the fourth and fifth positions, immediate recall of the codes would be better than when the figures were in any other positions. An experiment was carried out which tested recall with the pair of figures in each of the five possible locations. The prediction was found to have been correct, and the underlying reasons for making it to have been substantially supported. (author)

4/65–218 **Record Identification Using Variable Alphanumeric Names.** D. G. W. Thomas. CMPJ (Gt. Brit.), vol. 7, no. 1. April 1964. pp. 19–22.

This paper considers the problem of using a variable-length alphanumeric name, which may be read into a computer system in a variety of forms, as a key or file reference, and describes a data input program which was written for the English Electric-Leo KDP 10 which successfully employed this technique. (author)

4/65–219 **A Working System for Retrieval of Chemical Structures Adaptable to Pesticidal Screening Data.** Joe R. Willard and Edward J. Malkiewich. JCHD, vol. 4, no. 4, October 1964, pp. 211–217.

Describes in detail the coding of the IBM card.

4/65–220. **A Notation System for Indexing Pesticides.** Herman Skolnik and Avis Clow. JCHD, vol. 4, no. 4, October 1964, pp. 221–227.

Describes the basic symbols and rules for indexing in the Hercules notation system.

4/65–221 **The Atom Connectivity Matrix (ACM) and Its Characteristic Polynomial (ACMCP).** Leonard Spialter. JCHD, vol. 4, no. 4, October 1964, pp. 261–269.

4/65–222 **A Linearization of Chemical Graphs.** H. Hiz. JCHD, vol. 4, no. 3, July 1964, pp. 173–180.

The notational devices described are a development of the so-called Polish notation for logic and methematics, but are more involved because the chemical graphs are not exclusively trees.

4/65–223 **Generic (or Markush) Groups in Notation and Search Programs, with Particular References to Patents.** G. Malcolm Dyson. ISSR. vol. 2, no. 2, July 1964, pp. 59–71.

4/65–224 **A Polish-Type Notation for Chemical Structures.** Sylvan H. Eisman. JCHD, vol. 4, no. 3, July 1964, pp. 186–190.

The notation system described here is an atom-by-atom representation of a chemical compound, rather than a cipher or a fragmentation code. It is designed primarily for use within an information machine and particularly to reduce the amount of storage space required for the file of structural formulas.

Special Processing Languages

4/65–225 **Advanced Language Processing Procedures.** September 1963. S. J. Keyser. Electronic Systems Div., Air Force Systems Command, Bedford, Mass. ESD TDR63 620. (AD–430 608). 24 pp.; OTS price $2.60.

A recent theoretical advance is outlined in the science of linguistics which renders possible the use of natural English as a query language in computer based systems. The advantages of this approach are considered. (author)

4/65–226 **Some Problems in Information Science with Emphasis on Adaptation to Use through Man-Machine Interaction.** Final rept. vol. 1. January 1-December 31, 1963. Manfred Kochen. Thomas J. Watson Research Center, Yorktown Heights, N. Y. Contract AF19 628 2752, April 2, 1964. Proj. 5632, Task 563205 AFCRL 64 87 (AD–600 113). 213 pp.; OTS price $3.50. Vol. 2 is AD–600 047. OTS price $3.00.

An adaptive system for directly recording and retrieving information in simple, formal, English-like sentences; A description of an experimental information center and the language Query; Total information systems in planning and alerting; Working system for research in simulation of concept learning; Toward information systems science: information flow patterns and self-regulating mechanisms in the natural settings of libraries; Preliminary operational analysis of a computer-based, on-demand document retrieval system using coordinate indexing; Systems of orientation: a preliminary conceptualization; A model for the process of learning to comprehend; Methodology for research in concept-learning; Construction of a growing thesaurus by conversational interaction in a man-machine system; and Evaluation of bias in the removal of edges from a completely connected graph with an application to comparison of thesauri.

4/65–227 **An Analysis of Questions.** Preliminary rept. June 3. 1963. Nuel D. Blemap, Jr. System Development Corp., Santa Monica, Calif. Technical memo. no. 1287 000 00 (AD–422 845). 160 pp.; OTS price $12.00.

Analyzes the logical syntax of questions and answers, and sets forth the principle that a question is wholly determined by specifying the alternatives it presents and the request it makes. Shows how this analysis leads to a cross-classification of questions into several interesting varieties. Lists and discusses three varieties of questions. Describes a formal language (L) for which a logic of questions is devised by the addition of some machinery for asking questions in an orderly and fruitful way. Discusses the criteria which the formalization of questions and answers should satisfy. Outlines considerations of the notion of *truth* and *falsity* as these are related to questions. Sets forth semantical concepts in the appraisal of questions and answers, and presents a set theoretical analysis of questions and answers. (author)

4/65–228 **Information Grouping Logic—A Man/Machine Information Retrieval Communications Technique.** Delbert L. Ballard, John A. Liles and James S. Wood. *Parameters of Information Science.* pp. 285–295. (see 4/65–1)

Information Grouping Logic (IGL) is a notation which grew out of attempts by the authors to devise a computer procedure for retrieving from an inverted index file the *best available answer* to an interrogation against the file. IGL embraces all of the conventional Boolean operations as well as the exclusive OR. However, the notation lends itself readily to the simple and straightforward expression of interrogations of any order of complexity. The IGL notation that ultimately evolved is perfectly general and self-contained, and independent of the type of information to be handled. It can be made sensitive to the magnitude of numerical items, to the ranking of items by their importance, or to the sequence of items. IGL is described in abstract symbolic form and compared with conventional logic. The application of IGL is illustrated by typical examples. (authors)

4/65–229 **A Programming Language for Documentation—DOL.** Mandalay Grems. *Parameters of Information Science.* pp. 297–301. (see 4/65–1)

This paper describes a programming language for documentation that is mainly descriptive and is similar to COBOL. It provides the means for: a technical writer to prescribe instructions that establish the format, style and selected contents of a document; a typist to record these instructions along with the text, where the text includes both the words and the approximate punctuation; a proofreader to insert instructions for corrections to the document, and for composition of the document (typesetting). (author)

4/65–230 **A Comparison of List-Processing Computer Languages, Including a Detailed Comparison of COMIT, IPL-V, LISP 1.5, and SLIP.** Daniel G. Bobrow and Bertram Raphael. CACM, vol. 7. no. 4. April 1964. pp. 231–240.

A detailed comparison is presented of COMIT, IPL-V, LISP 1.5 and SLIP—four well-known computer programming languages which, among them, exhibit all the principal characteristics of existing list-processing languages. Important common features of list-processing languages are reviewed: forms of data structures which are manipulated, necessity for dynamic allocation of storage, use of pushdown stores, and use of recursive operations.

Principal differences between the four languages under consideration are detailed: representations of data, both by the programmer and within the machine; methods for storage allocation; programming formalisms and special processes available, including arithmetic facilities; and usability in terms of availability, documentation, learning aids and debugging facilities. A rough comparison shows that all the languages discussed have approximately the same speed.

Finally, the authors give some heuristics to aid in the selection of one of these languages for use in particular problem applications, concluding that no one of the languages considered is distinctly superior in all possible list-processing applications. (authors)

4/65–231 **On the Equivalence of Models of Language Used in the Fields of Mechanical Translation and Information Retrieval.** Maurice Gross. ISSR, vol. 2, no. 1, April 1964. pp. 43–57.

The notion of equivalence is defined formally; the model to which the other models are equivalent is Chomsky's model of context-free languages.

Equipment and Material Descriptions

4/65–232 **Unit Record Screening Viewer.** Final rept. June 1963. C. S. Axtell. Bausch and Lomb Inc., Rochester, N. Y. Contract AF30 602 6612, Proj. 6244, Task 624412 (AD–419 914). 12 pp.; OTS price $1.60.

The Unit Record Screening Viewer is a high resolution, rear projection viewer designed primarily for screening large numbers of mounted 70 x 100 millimeter film chips. The photographic images and the human readable code recorded on the film chip are projected at a fixed 6.5× magnification on the integral screen. The motorized chip changer mechanism may be operated from either local or remote positions for sequential screening of up to 64 mounted chips when loaded in an appropriate magazine. Chip placement in the projection focal plane is accomplished by the use of an electro-mechanical shaft positioner containing a two position stop cam and pawl device. The two cam positions correspond to the Load and Operate positions of the instrument. The large projection format required a novel condenser system utilizing rectangular shaped lenses allowing for space considerations. The film gate temperture is held within reasonable limits by the incorporation of a heat controlling multi-layer filter in the condenser system. A major problem was concerned with the magazine advance requirements. The mount thickness and magazine slot dimension decreed that the magazine must be advanced one step plus or minus 0.003 inches to permit the mount to re-engage properly in the magazine. Both the mount and the slot had square edges thus requiring precise indexing. (author)

4/65–233 **Automatic Unit-Record Storage and Retrieval Device BS-6A.** Final technical rept. April 3, 1964. S. Spengler and L. Maisner. Houston Fearless Corp., Los Angeles, Calif. Rept. no. R113 64. Contract AF30 602 2553, Proj. 4594, Task 459402 RADC TDR63 503. (AD–435 465). 45 pp.; OTS price $4.60.

Project SCRAM's prime program objective is to establish economical methods of storing 10,000 unit records and retrieving them at a minimum rate of 150 per second per module. Hydraulic and pneumatic methods of unit-record transport proved operationally unreliable and difficult to implement during subsequent breadboarding, so they were abandoned. More reliable and economical electromechanical hardware was developed. Logic subsystem design, special memory-control circuit design, design of commercial logic cards, and the drum memory design are discussed in this report. (author)

4/65–234 **Photocopying from Bound Volumes, Supplement No. 3.** William R. Hawken. Chicago, Library Technology Project, American Library Association, 1964. 40 pp. Ill. Paper. Price $5.00.

Evaluated in this report are the Minnesota Mining and Manufacturing Company's automatic Model 209 Dry Photo-Copier, Model 76 Dry Photo-Copier, and De Luxe Transparency Maker Model 70; the Pacer International Corporation's Sight-scope Exposing Unit and Star Photocopier, and the Victoreen Instrument Company's Vico-Matic Copier. For each machine the report gives a general account and specifications, an analysis, a description of techniques of use, cost studies, and an estimate of performance. Pictures—twenty in all—supplement the text. Format of the report has been planned for compatibility with *Library Technology Reports*, the Library Technology Project looseleaf subscription service due to begin in January 1965. The looseleaf format permits the replacement of a report on an item of equipment with a new report from time to time. (Council on Library Resources)

4/65–235 **Factors Influencing the Design of Original-Document Scanners for Input to Computers.** August 19, 1964. Edward S. Stein and Associates. Tech. note 245, National Bureau of Standards, Washington, D. C. USGPO price $.35.

This report considers some of the factors involved in the design and implementation of scanners capable of reading data from documents in their original form rather than from microfilm. The scanner is assumed to be part of a transcribing information processor such as the FOSDIC (*F*ilm *O*ptical *S*canning for *D*irect *I*nput to *C*omputers) type of machine. Advantages and disadvantages of microfilm as against original document scanning are discussed. Factors considered include the types of forms that are to be handled, paper handling problems, available electronic and mechanical scanning techniques, problems of resolution and image defects, precision limitations, and overall system limitations. A set of general specifications for scanners for direct data input to computers is developed in qualitative terms. (author)

4/65–236 **Associative Memory Algorithms and Their Cryogenic Implementation.** December 1963. J. L. Rogers and A. Wolinsky. Space Technology Labs., Inc., Redondo Beach, Calif. Rept. no. 8670 6007RU000. Contract Nonr3839 00. (AD–429 521). 113 pp.; OTS price $9.60.

The report has two aspects. One is the design of a versatile cryogenic associative memory; the other is the analytical survey of existing algorithms for associative memories, the description of certain new algorithms, and the discussion of the implementation of these algorithms in the associative memory of this report. The cryogenic associative memory described herein gains its versatility from the presence of a small number of tag bits of special design in each word, which may be used for the marking of words which satisfy certain criteria. In the discussion of associative memory algorithms, new variations of algorithms are developed in several instances which are in various ways superior to the existing ones. The implementation of the algorithms in the general-purpose associative memory of this report is also discussed and a number of new applications of associative memory techniques are described. (author)

4/65–237 **Summary of Investigation on Associative Memories.** January 15, 1964. Computer Command and Control Co., Philadelphia, Pa. Rept. no. 5 101 5. Contract Nonr4068 00. (AD–428 577). 11.; OTS price $1.60.

4/65–238 **Research on Ferret Associative Memory.** Interim engineering rept. April 15–April 30, 1964. Joseph Caschera. Philco Corp., Willow Grove, Pa. August 1964, 1 vol. Contract AF33 615 1259, Proj. 4028, Task 402830. (AD–445 796).

This report describes the design and development of a 1024 word, 48 bit associative memory capable of performing exact match searches with completely variable masking as well as operation as a conventional nondestructive readout random access memory. The interface is, with the exception of a more elaborate control register, the same as that of a conventional memory, thus allowing greater flexibility in its application to computing systems. The system uses Biax ferrite elements for storage, Micrologic integrated circuits for control and timing logic, and discrete components circuits for current pulse driving. The memory is divided into two groups of 512 words for sharing of the detectors and detector amplifiers. A search on each bank requires six microseconds and the addresses of the solutions may be read out at a one megacycle rate. Normal interrogate (read) is nondestructive and can be randomly accessed at a one megacycle rate. A description of the system, circuits, and operating characteristics is presented. (author)

4/65–239 **The Logical Design of a Multichannel Device for the Retrieval of Information.** April 1964. Vance R. Wanner. Office of Naval Research, Washington, D. C. ONR ACR93. (AD–601 987). 208 pp.; OTS price $3.50.

4/65–240 **A Discussion of Associative Memories from a Device Point of View.** A. E. Slade. *Parameters of Information Science.* pp. 331–335. (see 4/65–1)

During the past few years a number of memories have been proposed where information is retrieved on the basis of its content rather than its address. Part of the reason for increased emphasis on associative memories is that the cryotron has made construction of large stores of this type seem feasible. This paper will describe various associative memory designs and discuss the present state of the art. (author)

4/65–241 **A New Optical Marking and Search Technique.** W. H. Libby. *Parameters of Information Science.* pp. 307–310. (see 4/65–1)

A new technique for marking a card-like information storage medium with desired search designations has been developed. It involves the generation within a transparent card, of light-reflecting marks which are searched by illumination from the edge of the card, permitting parallel search and making available the entire card area for encoding. Since the marks do not affect light directed normally to the card, the entire surface may be used for transparent or opaque imaging. (author)

4/65–242 **A Comparative Analysis of Photocopying Processes.** David C. Kronick. *Reproduction Methods for Business & Industry.* Vol. 4, no. 12. December 1964. pp. 52, 68.

This brief excerpt from the author's MBA thesis gives results of a survey of electrostatic, thermopgarphic, diffusion-transfer, dye-transfer, and diazo reproduction machines. The most interesting data shown pertain to percentage user satisfaction, percentage waste, volume of use, and best and least liked machine features. (C.L.B.)

4/65–243 **Microscopic Blemishes and the Storage of Permanent Record Microfilm in Perspective.** June 1964. Vernon D. Tate. NAMI No. 70. pp. 242–248.

4/65–244 **Microscopic Spots: A Progress Report.** June 1964. D. G. Wiest and R. W. Henn. NAMN No. 70. pp. 249–276.

4/65–245 **Photochromic Information Storage.** Joseph Becker. DTMN. vol 10, no. 12. December 1964. pp. 32–33.

4/65–246 **The Mechanized Library.** L. H. Martin. DTMN, vol. 10, no. 9. September 1964, pp. 32–36 (a description of the AVCO Verac photographic storage system).

4/65–247 **Developments in Document Reproduction.** April 28-29, 1964. Loretta J. Kiersky. SPLB, vol. 55, no. 6, July-August 1964, pp. 388–389.

Describes some of the equipment displayed at the 13th Annual Meeting and Convention of the National Microfilm Association, in Philadelphia.

Use and User Studies

4/65–248 **The Utilization of Information Sources During R & D Proposal Preparation.** October 1964. Thomas J. Allen. Report 97–64 of the Alfred P. Sloan School of Management. M.I.T., Cambridge, Mass. 31 pp.

Twenty-two proposal competitions for government R & D contract, involving 156 proposal teams, are examined to determine the relative use of three sources of technical information. The extent to which each proposal team relied upon literature search, the use of staff specialists within the lab and the use of outside sources of information is related to the rated technical quality of its proposal, and to other variables characterizing the proposal team and its parent laboratory.

Twenty-two percent of the total time expended by 156 proposal teams was devoted to the seeking and gathering of technical information. Of the three information sources used, only one, laboratory specialists, appears to be at all directly related to the technical quality of the product and this relation is weak and unreliable. Technical quality is inversely related to the extent to which the proposal team relies upon individuals outside of the laboratory as sources of information. (author)

4/65–249 **Environmental Information Needs of an Army RDT&E Laboratory.** April 1963. David Askin. Army Frankford Arsenal, Philadelphia, Pa. Proj. 1A025001A622 AFA A63 17 (AD–430 579). 22 pp.

This paper outlines the needs of the designer, the reliability engineer, the product improvement redesign engineer, the quality control engineer, the specification writer, and the environmental test engineer. It describes available sources for such information, and their limitations. It points out in what form this environmental information should be presented, and recommends means for establishing more effective communications in this field. (author)

4/65–250 **Activity Statistics for a Large Bio-Medical Library.** Part 2 of the final report on the Organization of Large Files. April 30, 1964. Advanced Information Systems Division of Hughes Dynamics, Inc., Sherman Oaks, Calif. 20 pp.

The study of library holdings and circulation statistics was undertaken to clarify the nature of the various distributions which were defined in the prior study. The prior study presented the use of a measure of efficiency so as to optimize the file operation in terms of the physical arrangement and showed that this involved an organization based upon activity on the file. The distributions defining this activity are: the variable concentration of file items which may be considered as reflecting past file activity; the worth of file items related to current activity; and the probable future activity on the file items. The results of the current study provided in this report indicate that a large degree of predictability is inherent in the file operations and that this predictability can be utilized to provide efficient operation. (author)

4/65–251 **Report of an Investigation on Literature Searching by Research Scientists.** John Martyn. 1964. Aslib. London, England. 20 pp. price 7*s.* 6*d.*

4/65–252 **Search Strategy by Reference Librarians.** Part 3 of the Final Report on the Organization of Large Files. March 17, 1964. G. Carlson. Advanced Information Systems Division, Hughes Dynamics, Inc., Sherman Oaks, Calif. 40 pp.

A study was made to describe in detail the search procedures used by reference librarians. It was expected that the results of this study would illustrate that human searching behavior can be precisely described. By highlighting human inconsistencies and ambiguities, it is hoped that the need for improvements in the manual system of searching will become evident. Solutions to the deficiencies of manual search lie in new training procedures and automated assistance.

The basic procedure was to present a search problem to a reference librarian and then record in detail his search techniques. An heuristic analysis was made of the observations. The result of the analysis is a flow chart with almost enough precision so that a computer program could be written directly from the charts. The flow charts lack accurate numerical values for some parameters and do not account for procedures which may have been omitted by the librarians. (author)

4/65–253 **Interview Guide Handbook.** May 8, 1964. H. F. Sieber, Jr. Auerbach Corp., Philadelphia, Pa. Contract SD219 (AD–601 259). 51 pp.; OTS price $1.50.

One of the principal tasks of the Department of Defense is research and development leading to the production of particular weapons systems. These research and development (R and D) efforts involve searching for and utilizing astronomical amounts of scientific information. It has been estimated that scientific knowledge is doubling every ten years. As a result, increasing amounts of research and development effort are wasted in searching through these growing mountains of information. The problem, therefore, is to channel the pertinent information to interested personnel as efficiently as possible. But, in order to devise a system for more efficient delivery of this information, there must be insight into the information use patterns of these personnel. The aim of this study is to provide that insight, that is, to determine how Department of Defense scientists and engineers obtain and use scientific and technical information. For this study, the AUERBACH Corporation has developed an interview guide for use in interviewing approximately 1,500 personnel randomly selected from the DOD Research, Development, Test and Evaluation (RDT and E) population. The interviews average about two hours in duration and will be semi-structured in nature. (author)

4/65–254 **A Preliminary Study of Information Exchange Activities of Foreign Psychologists and a Comparison of Such Activities with Those Occurring in the United States.** June 1964. Amer. Psychological Assoc., Washington, D. C. APA–PSIEP–10. NSF Grant GN–281. 38 pp.

4/65–255 **The Discovery and Dissemination of Scientific Information Among Psychologists in Two Research Environments.** Report APA–PSIEP–11 of the American Psychological Association, Washington, D. C. September 1964. NSF Grant GN–281. 125 pp.

The general objective of this study was to develop a description of the scientific information exchange environment in psychology by determining the responsibilities of the persons and institutions associated with the origination, transmission, storage, and use of scientific information; the characteristics of the processes involved in the fulfillment of each of these functions and of the storage properties they afford; and, finally, the utilization of scientific information. To accomplish this objective sixteen specific studies were designed of which nine are available as technical reports. Findings trace the dissemination of information from initiation of research through archival publication (stressing that type of communication occurring at conventions); discuss access to and use of archival media; and determine the characteristics of the user's behavior in his search for and use of needed information. Subsequent work will emphasize the user's behavior and will develop and test communication innovations which seem practical and desirable. (authors)

4/65–256 **Systems in Scientific Information Exchange and the Effects of Innovation and Change.** Belver C. Griffith and William D. Garvey. *Parameters of Information Science.* pp. 191–200. (see 4/65–1)

Possible innovations in disseminating, storing and retrieving research reports are viewed against some data and theory on scientific information exchange. While certain innovations may simply ameliorate particular problems, others may have far reaching effects; even extending into the conduct of scientific work or the organization of the science into subareas. (authors)

4/65–257 **The Effect of Convention Presentations on Information Exchange Behavior and Subsequent Research.** William D. Garvey and Belver C. Griffith. *Parameters of Information Science.* pp. 201–213. (see 4/65–1)

The paper describes an effort to study intensively the behavior of scientists within a specific scientific information-exchange environment, the annual convention. For the scientists peopling this environment, an attempt is made to determine the effects the convention presentations have on their scientific activities after the convention. (authors)

4/65–258 **The User and the Technical Information Center.** C. Allen Merritt. *Parameters of Information Science.* pp. 255–258. (see 4/65–1)

Providing timely, accurate, and complete information to its users is the primary objective of a technical information center—"getting the right information to the right person at the right time." How a system does this is of no direct interest to the user; his main concern is with results. If his individual needs are not met adequately, he immediately concludes that the information center is of no value. To avoid this unfortunate reaction requires constant monitoring of what a user gets from a system and how valuable it is to him. This paper describes computer programs that analyze and evaluate the output of a technical information center, to achieve a high degree of precision and relevancy in disseminating information to the user. (author)

4/65–259 **The "Information Explosion" and the User's Need for Hard Copy.** A. Resnick. *Parameters of Information Science.* pp. 315–317. (see 4/65–1)

An experiment was conducted in which the availability of hard copies was systematically restricted in IBM's Selective Dissemination of Information (SDI) System. The purpose of this study was to determine the extent to which users of technical information would make an effort to obtain copies of documents which were of interest but were no longer easily obtained. The results support previous findings as depicted in Mooers' and Zipf's Laws. Users made significantly less effort to obtain documents of interest for those documents which were deliberately withheld than for those documents which were more readily available. (author)

4/65–260 **The Methodology of the DOD Scientific and Technical Information Use Study.** H. F. Sieber. *Parameters of Information Science.* pp. 235–242. (see 4/65–1)

This paper describes the objectives and scope of the Department of Defense (DOD)-sponsored study of how research, development, test, and evaluation (RDT&E) personnel of the Army, Navy, and Air Force require and use scientific and technical information. It includes a detailed discussion of the study methodology, survey techniques and plans for analysis of the results of the survey. (author)

4/65–261 **The Structure, Objectives, and Findings of a Study of Scientific Information Exchange in Psychology.** William D. Garvey and Belver C. Griffith. AMDO vol. 15, no. 4. October 1964, pp. 258–267.

4/65–262 **Information Needs of the Technical Organization.** George A. Whittington. REDE, vol. 15, May 1964. pp. 18–21.

Discusses eight steps as an individual or group solution to problems involved in meeting information needs.

4/65–263 **Unintentional Duplication of Research.** John Martyn. NNSC, vol. 21, February 6, 1964. p. 338.

A survey conducted by the Aslib Research Department reveals instances of belated discovery of information in the literature, and estimates the cost of research duplication.

Design, Testing, and Evaluation of Information Systems

4/65–264 **The Evaluation of Automatic Retrieval Procedures—Selected Test Results Using the SMART System.** Gerard Salton, A Report of Computational Laboratory of Harvard University, Cambridge, Massachusetts. 1964. NSF Grant GN–245. 38 pp.

The present report deals with the evaluation of a variety of automatic indexing and retrieval procedures incorporated into the SMART automatic document retrieval system. The design of the SMART system is first briefly reviewed. The document file, search requests, and other parameters affecting the evaluation system are then examined in detail, following by a description of the measures used to assess the effectiveness of the retrieval performance. The main test results are given and tentative conclusions are reached concerning the design of fully automatic information systems. (author)

4/65–265 **An Analysis of Duplication and Consistency of Subject Indexing Involved in Report Handling at the Office of Technical Services.** U.S. Department of Commerce. March 1963. Ann F. Painter. Rutgers Univ. PB–181 501. 135 pp.; OTS price $2.75. Submitted as a thesis to the Graduate School of Rutgers. (see 4/64–164)

4/65–266 **The Scope and Operating Efficiency of Information Centers.** 1963. Richard M. Dougherty. The Chemical-Biological Coordination Center of the National Research Council. New Brunswick, N. J. 242 1. Microfilm. Thesis—Rutgers Univ.

4/65–267 **Methodology for Test & Evaluation of Information Retrieval Systems.** July 1964. W. Goffman and V. A. Newill. Report CSL:TR–2 of the Center for Documentation & Communications Research. Western Reserve University, Cleveland, Ohio.

4/65–268 **A Use for the Techniques of Structural Linguistics in Documentation Research.** September 14-18, 1964. Jessica S. Melton. Report CSL:TR–4 of the Center for Documentation and Communication Research. Western Reserve University, Cleveland 6, Ohio. A paper presented at the International Study Conference on Classification Research. Elsinore, Denmark. 20 pp.

Index language is considered as a separate component and a variable in an information retrieval system. It is suggested that for purposes of testing, comparing and evaluating index language, the techniques of structural linguistics may provide a descriptive methodology by which all such languages (hierarchical and faceted classification. analytico-synthetic indexing, traditional subject indexing, indexes and classifications based on automatic text analysis, etc.) could be described and compared on a common basis.

4/65–269 **Centralization and Documentation.** 2nd ed. June 1964. Report C–64469 of Arthur D. Little, Inc., Cambridge, Mass. 49 pp. An edited and revised edition of the original report (see 4/64–14), with some new material incorporated to extend the original model.

4/65–270 **An Evaluation of Links and Roles Used in Information Retrieval.** July 1964. Jefferson D. Sinnett. A. F. Materials Lab. Research and Technology Div., A.F. Systems Command. Wright-Patterson Air Force Base, Ohic. Tech. Doc. Rep. ML–TDR–64–152. 139 pp.

Links and roles are techniques for modifying a basic coordinate indexing method. The purpose of the modification is to reduce the amount of irrelevant information retrieved without losing relevant information. Theory of retrieval is reviewed, mathematical models are discussed, retrieval is described in terms of basic concepts for algebra of sets, and a method of scoring effectiveness is presented. Roles were found inferior to links. Links showed reduction of irrelevant information by over 56% while incurring less than 5% loss of relevant information. (author)

4/65–271 **Methodologies for System Design.** Final rept. February 24, 1964. Hughes Dynamics, Los Angeles, Calif. Contract AF30 602 2620. Proj. 4594. Task 459 405. RADC TDR 486. (AD–434 749). 101 pp.; OTS price $9.10.

This report presents the results of work on development, programming, and testing of methodologies for aiding information system design and evaluation. They include, 1. a first step in formalization of a calculus of operations developed for aiding definition of processes for file organization and searching; 2. the complete programs for an evaluation and assignment model which provide for mechanized determination of the optimum assignment of components and functions to points in a hierarchical reporting structure, 3. test results on the relative effectiveness (in terms of quality of results and time and cost for the design process) of the measure of system efficiency, the system design model, and the evaluation and assignment programs. (author)

4/65–272 **Methodologies for System Design.** Appendices to final rept. February 24, 1964. Hughes Dynamics, Inc., Los Angeles, Calif. Contract AF30 602 2620. Proj. 4594. Task 459 405. RADC TDR63 486A (AD–434 748). 237 pp.; OTS price $15.50.

4/65–273 **Final Report on the ROUT Document Retrieval System.** May 1964. J. F. Rial. The MITRE Corp., Bedford, Mass. Contract AF 19 (628)–2390. ESD-TDR–64–96. TM–3869. AD–601 145. 66 pp.; OTS price $1.75.

The evaluation of the ROUT (Retrieval of Unformatted Text) document retrieval system consisted of a detailed examination of various retrieval errors in the context of different query-processing methods. Many of the methods used to extend the power of the basic coordinate indexing mechanism were compared. Three different evaluation techniques were used in the study. Based on the analysis, conclusions were drawn about the relative merits of several query-processing methods, and some general comments were made about the limitations of coordinate indexing as applied to the document retrieval problem. (author)

4/65–274 **Progress Report on an Operations Research Systems Engineering Study of a University Library** (National Science Foundation Grant NSF–GN–31). Principal investigator: Robert H. Roy, Baltimore, Md. Johns Hopkins University, 1963. 144 pp.

4/65–275 **Applications of Serial and Parallel File Structures to Command and Control Systems.** August 1964. Stanley B. Rosen, John P. McCabe, and Essie S. Nevans. General Electric Co., Washington, D. C. Contract AF19 628 3843, Proj. 2801, Task 280101. ESD TDR64 519 (AD–604 862). 46 pp.; OTS prices: HC $2.00/MF $.50.

This study was undertaken to determine if the introduction of redundant files can reduce search time when serial outputs are required. Serial and parallel file structures on magnetic tape and disc memories were studied. The conclusion was made that search time can be improved because of the inherent random-access features of disc memory. Further studies relevant to the study are suggested. (author)

4/65–276 **Purpose and Objectives of the Comparative Systems Laboratory.** A. J. Goldwyn. August 1964. Report CSL: TR–1 of the Center for Documentation & Communication Research. Western Reserve University, Cleveland, Ohio.

Modern documentation studies concentrate on system analysis and evaluation, research in the past and present nature of scientific communication, and training. The Comparative Systems Laboratory (CSL) at Western Reserve is planned to isolate and to test the component parts of an IR system; to study the concept of relevance; to collect information on the medical vocabulary (-ies); and to investigate certain areas of automatic language processing. Experimental design has included defining components and establishing measures for evaluation. Close liaison with potential user groups will provide "real" questions and evaluation for the actual testing phase. The articulation of a training program with the CSL will enhance its long-term contribution to the research community.

4/65–277 **File Reference.** August 18, 1955. J. A. Postley. P–691. RAND Corp., Santa Monica, Calif. (AD–604 720). 22 pp.; OTS prices: HC $1.00/MF $.50.

Suggestions which will be useful in the problem area of reference to large files of information are presented. While the problem of indexing is essentially the same in any file reference problem and equipment, it can always be solved by keeping the (exhaustive) index—not the file—up to date. In some cases it is possible to appreciably reduce the size of the index which may be required. Undue emphasis on the factor of pipeline time may be removed in favor of increased emphasis on that of traffic rate. This may be done with the same equipment in three different ways, namely: 1. by approximating the optimum balance of emphasis within a single problem; 2. by emphasizing one of these factors in one problem and another in another problem, within the same framework of mechanization; and 3. by emphasizing one factor, e.g., traffic rate, throughout the problem and handling requests requiring an emphasis of pipeline time reduction as special cases, emphasizing this factor. (author)

4/65–278 **The Organization of Large Files—Introduction and Summary.** Part 1 of the Final Report on the Organization of Large Files. April 30, 1964. Advanced Information Systems Division, Hughes Dynamics, Inc., Sherman Oaks, Calif. 23 pp.

The final report on the study of the organization of large files has been prepared in the form of six separate reports, to which this is the introduction and summary. The study investigated various relationships between logical aspects of information retrieval (such as item representation and relevancy) and physical aspects (such as indexing structures and procedures for reorganization). Specifically, the study included: 1. operating library statistisc on activity; 2. search strategy protocols of reference librarians; 3. procedures for file reorganization; 4. procedures for word association and file conversion; and 5. the measurement of file information and associated file organizations.

The statistical results confirmed the extent to which activity organization can be applied to operating libraries. The search strategy protocols demonstrated the extent to which reference librarians follow relatively well-defined heuristics, which can be applied not only to mechanical searching but to training of library personnel as well. The procedures for file reorganization have been programmed and tested sufficiently to demonstrate effective capability for adjustment to environment. The word association concepts have been flow charted and conversion procedures developed for application to the tape files of the Defense Documentation

Center. The mathematics for measuring information from a file, particularly in terms of its relationship to relevancy, error, and index structure, has been defined. (author)

4/65–279 **Self-Organizing Processes—Computer Programs for Files of Bit Patterns.** Part 4 of the Final Report on the Organization of Large Files. April 30, 1964. Ralph M. Shoffner. Advanced Information Systems Division of Hughes Dynamics, Inc., Sherman Oaks, Calif.

To provide an experimental vehicle for testing various procedures for file organization, various criteria for logical relationship, and various measures of performance, a *bit-processor* has been programmed which organizes a file of patterns of bits so as to automatically develop indexing structures which reflect skewed distributions of activity. The programs are designed to allow various subroutines, criteria, and measures to be changed. The programs have been debugged and run sufficiently to demonstrate that they do in fact organize the file so as to be responsive to activity. (author)

4/65–280 **Self-Organizing Processes—Programs for Files of Word Associations and for Conversion of Document Descriptions.** Part 5 of the Final Report on the Organization of Large Files. April 30, 1964. Advanced Information Systems Division, Hughes Dynamics, Inc., Sherman Oaks, Calif. 64 pp.

To provide a vehicle for testing self-organizing procedures which would have some intuitive significance, a word association model has been developed. This model is described and the flow charts for its operation presented. Flow charts for the conversion of the tape files of the Defense Documentation Center are also presented. (author)

4/65–281 **Measurement of File Operating Effectiveness—Time, Cost, and Information.** Part 6 of the Final Report on the Organization of Large Files. April 30, 1964. R. M. Hayes. Advanced Information Systems Division, Hughes Dynamics, Inc., Sherman Oaks, Calif. 50 pp.

A criterion for evaluation of a file organization, suitable for large files, is developed, based on response time, T, cost, C, and a measure of *information*, I. The usual entropy measure for information is extended to include the relevancy of the information received as a significant parameter. The effect of errors in file operation is considered, and it is shown that these are best combatted by responding with a set of items rather than a single one. This is applied particularly in relationship to search strategy in an indexed file, in this way extending the methods of sequential decoding to file operation. To reduce the effective searching time, the techniques of activity organization can be used; these are quantitatively characterized. (author)

4/65–282 **The Organization of Files for Command and Control.** Interim technical rept. March 25, 1964. N. S. Prywes. Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pennsylvania. MSEE 64 12. Contract Nonr551 40 (AD–600 980). 38 pp.; OTS price $1.00.

This paper considers the problem of maintaining large files describing complex and widespread operations in a computer as well as storing and retrieving the necessary information in real time. The requirements are first stated generally, in terms of the hierarchy of the organization, hierarchy of the computers used, the language of communication with the file and the retrieval and updating. The paper then turns to a more detailed description of specific features and techniques. Three types of files are considered: Resources File, Plans File, and a Technical Library File. The organization of each file, use of trees and lists and loading of files are described. Techniques are given for retrieval by position in an organizational hierarchy, cross referencing between files, and the creation of a thesaurus and a classification system. (author)

4/65–283 **Optimum Storage of Library Material.** 1964. Julius Grady Cox. Purdue University Libraries, Lafayette, Indiana. 222 pp. (see 4/65–284)

Thesis accepted by the faculty, Purdue University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy, June 1964, accomplished under a National Science Foundation Faculty Fellowship. Three constrained models were developed and tested against Purdue and Auburn University libraries. Only a small number of shelf heights, from 3 to 5, was required to provide nearly an optimal solution. (C.D.G.)

4/65–284 **Compact Book Storage in Libraries.** Ferdinand F. Leimkuhler and J. Grady Cox. OPRE, vol. 12, no. 3. May–June 1964, pp. 419–427.

A method is developed for optimally shelving inventory items by size, with particular reference to large library collections. The area presented by $n + 1$ distinct shelf heights is minimized for any collection that can be characterized by an item-height distribution. When item-height is continuous, the necessary conditions for the extremum are recursive in the optimal shelf heights, and the solution reduces to a simple computational search. For certain distribution functions, as demonstrated in the example, an additional recursive relation between the solutions for different n further simplified the computation considerably. Geometrical representations of both the model and the solution method are presented. The maximum increase in storage capacity can be expressed generally as a simple function of the mean and the maximum item-height. (authors)

4/65–285 **A Statistical Optimization of Search Time in an Information Retrieval System.** Interim technical rept. January 1964. Philip Leslie Leifer. Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa. MSEE64 17. Contract Nonr551 40 (AD–600 978). 63 pp.; OTS price $1.75.

The Magnacard system is used to store large files of inventory information. It consists of a bank of filing cabinet drawers containing oxide coated cards, each one inch by three inches, which can be read from and written onto by magnetic heads similar to those found in tape recorders. Access to the stored information is provided through a mechanism which can remove a drawer from the bank, read from or write onto the cards in that drawer, and return the drawer to the bank. Consideration is given to the organization of the inventory file within the drawers, in order to optimize the retrieval time for a given number of interrogations a of the file. For a given size f and a given number of interrogations a, in a time period T, there is a value d which will minimize the average time to satisfy the a requests. (author)

4/65–286 **On the Utility of the Releavance Concept.** March 18, 1964. Carlos A. Cuadra. System Development Corp., Santa Monica, Calif. Rept. no. SP1595 (AD–443 746). 9 pp.

This report argues that it is necessary to separate relevance as a construct from the particular means used to measure it. The notion (Bridgeman) of operation definitions can be very helpful in information system evaluation because it permits to avoid the nonproductive issue of defining 'needs themselves' to the more fundamental question of what are the utility and consequences of using particular measures of relevance. (author)

4/65–287 **Evaluation of the 1963 ADI Meeting.** 15 January, 1964. Carlos A. Cuadra. *ADI Proc.* Pt. 3. (see 4/65–1)

This paper reports the analysis of a questionnaire completed by 235 of the attendees at the 26th Annual Meeting of the American Documentation Institute, October 6–11, 1963, in Chicago. Attendance and evaluative data are presented and suggestions are reported for the improvement of the future meetings. The complete paper, published as SP–1514 at System Development Corporation, was submitted as the report of the ADI Evaluation Committee. (author)

4/65–288 **The Analysis of Index Language Devices.** 1964. C. W. Sleverdon and J. Mills. *ADI Proc.* Pt. 3. (see 4/65–1)

Systems design demands full awareness of all the working elements of the system. In the vital indexing process the fundamental constituent elements have never been thoroughly examined. These fundamental elements are the various devices for defining the classes to which documents are assigned and in terms of which all searching is made. Indexing systems (UDC, "Uniterm," etc.) are different aggregates of these indexing devices. Two groups of devices are distinguished: those which, if added to a raw vocabulary of single index terms, widen class definition e.g.,

coordination, addition of roles, links). A collection of 1,500 documents of high speed aerodynamics is being exhaustively indexed and 400 questions, against which each document has been assessed for relevance, will be used to derive performance figures (measured in recall and relevance) for each indexing device separately. (authors)

4/65–289 **Toward Document Retrieval Theory: Relevance-Recall Ratio for Text Containing One Specified Query Term.** 1964. Manfred Kochen, *ADI Proc.* Pt. 3. (see 4/65–1)

A trade-off relation between hit rate and acceptance rate is derived. Hit rate (acceptance rate) is defined as the expected number of documents retrieved and relevant, divided by the expected number of relevant (retrieved) documents. The derivation is based on the following assumptions. A document is retrieved if and only if it contains a given word explicitly in its title or text. The query word is judged relevant to the query topic by the querist. The probability that an author uses a word in an article, given that he judges the word irrelevant to the query topic, is expressed by the frequency-rank relation of the underlying language. The corresponding probability for relevant words is assumed to have the same shape with all ordinates raised. Acceptance rate is shown to decrease monotonically with hit rate for languages in which few words account for most of the use, to have a minimum for others. (author)

4/65–290 **Memory Organization of a 7090 to Do Statistical Association Processing for Document Retrieval.** M. E. Sherry. *Parameters of Information Science.* pp. 337–339. (see 4/65–1)

Statistical association processing of a large document collection for purposes of document retrieval is being accomplished through manipulation of numerical association matrices of dimension 1,000 and up, an order of magnitude larger than normally encountered in scientific data processing. Because these matrices are sparse, the storage of only non-zero values permits more efficient processing, particularly for matrix multiplication. A format for matrix storage on magnetic tape and in core memory, as well as a program for performing matrix multiplication, are described. (author)

4/65–291 **Project Lawsearch: A Statistical Comparison of Coordinate and Conventional Legal Indexing.** L. B. Heilprin and S. S. Crutchfield. *Parameters of Information Science.* pp. 215–234. (see 4/65–1)

An experiment was conducted to compare conventional with coordinate legal search. A given area of the law was indexed for coordination; and test questions in this area were answered by a group of searchers, by both methods. Some 200 cases were retrieved in roughly equal samples, and classified by applicability to the search questions. The classes were: 1. cases-in-point; 2. cases applicable by analogy; 3. dicta; 4. cases not applicable. A statistical analysis showed the two samples were roughly comparable as to applicability distribution; and that identical cases were found 100% in Class 1, about 15% in Class 2, 14% in Class 3, and 6% in Class 4. The probability of finding identical cases by individual searchers was much lower than for the samples collected by 4 or 5 searchers.

Analysis of total search times showed a mean conventional search time of 138 minutes, and a mean coordinate search time of 97 minutes. However, the mean coordinate search time decreased over the five days of tests—indicating a learning curve. A hypothesis was made that what was being learned was to invert the order of search customary in law—starting with a rather general class, searching it and rejecting non-applicable cases. This order was called general-to-specific search. With the coordinate system it was possible to search for the most specific class first, then broaden out the search if needed—the specific-to-general method. The time saved was attributed to the smaller amount of scanning required by the latter method. A test of this hypothesis was made and it was found that the more cases of highest applicability were present, the shorter the coordinate search relative to the conventional.

The implications of the inverse time-order approach in design of indexing systems are briefly discussed. (authors)

4/65–292 **A Comparison of Relevance Assessment by Three Types of Evaluator.** Gordon C. Barhydt. *Parameters of Information Science.* pp. 383–385. (see 4/65–1)

The effectiveness of measures of relevance based on non-user evaluation is investigated. Relevance assessments are made by three types of evaluator: user, expert, and system specialist. The test file consists of 4,000 educational research documents; the questions are those posed by a pilot user group (control). It is hypothesized that no significant difference exists among the relevance assessments of user, expert and system specialist for a set of responses to a given question. Results are given. (author)

4/65–293 **Basic Parameters of Retrieval Tests.** R. A. Fairthorne. *Parameters of Information Science.* pp. 343–345. (see 4/65–1)

Retrieval tests involve comparison of two partitions of the collection, by the system and by the tester. Parameters should hold good both for *all-but-not-only* (ABNO) and *only-but-not-all* (OBNA) systems; when there are no items of the type invoked; and when *discrimination* is separated from *identification.* Thus they are functions of those parameters of the 2 x 2 contingency table unaltered, except for sign, by row or column interchange. These are interpreted in documentary terms. (author)

4/65–294 **A Comparative Study of Fragment versus Document Retrieval.** J. L. Kuhns and C. A. Montgomery. *Parameters of Information Science.* pp. 369–377. (see 4/65–1)

4/65–295 **CITIS: A Generalized Mathematical Model for a Centralized Integrated Technical Information System, from a Resource Allocation Point of View.** Harold J. Podell. *Parameters of Information Science.* pp. 155–172. (see 4/65–1)

The CITIS (Centralized Integrated Technical Information System) models introduce a linear programming application for the information center manager. A successful test of CITIS Model I is conducted using the PRINCE (Parts Reliability INFormation CEnter) System, a NASA program. The CITIS Model I illustrates the key summary resource allocation relationships that are necessary for the effective development of a *centralized-decentralization* operation for a technical or a scientific information center. (author)

4/65–296 **Some Techniques for Measuring System Performance.** D. W. King and P. J. Terragno. *Parameters of Information Science.* pp. 393–398. (see 4/65–1)

Some emphasis has recently been directed to evaluating system performance in terms of a retrieval profile as described by relevant and non-relevant documents retrieved and not retrieved. Given one or more assessments of relevance it is possible to estimate these values using statistical sampling techniques. Examples are presented for both operating systems and systems in the developmental stages. (authors)

4/65–297 **An Experiment with File Ordering for Information Retrieval.** George Badger, Jr. *Parameters of Information Science.* pp. 379–381. (see 4/65–1)

4/65–298 **Testing, Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles.** Barbara A. Montague. *Parameters of Information Science.* pp. 357–367. (see 4/65–1)

This paper evaluates the effect of system components such as links, roles, depth of indexing, vocabulary control, generics, and search strategy on recall, relevance, and cost of a patent index. Input costs and retrievability effectiveness of a concept coordination IR system are compared with costs and effectiveness of a classification system. (author)

4/65–299 **Information Retrieval.** Martha E. Williams. *Frontier,* vol. 25, no. 3. Autumn 1964. pp. 10–13.

A brief, common-sense evaluation of man-machine IR systems, stressing the need to consider more than a single technique in catering to a specific user group. (C.L.B.)

LITERATURE NOTES—Please enter my subscription for one year commencing with the January 1965 issue. At $4.00 per year, payment is enclosed ☐

bill me. ☐

Name__________ Title__________

Address ☐ Business ☐ Home__________

City__________ State__________ Zip__________

Your Firm Name__________

Literature Notes Separate

The Literature Notes section of American Documentation is now available as a separate publication.

Literature Notes will continue to be an important section of American Documentation. However, popular demand has persuaded ADI to offer this separate quarterly subscription as an offprint.

On a Calendar year basis, the subscription price of Literature Notes is $4.00 per year, commencing with the January 1965 issue.

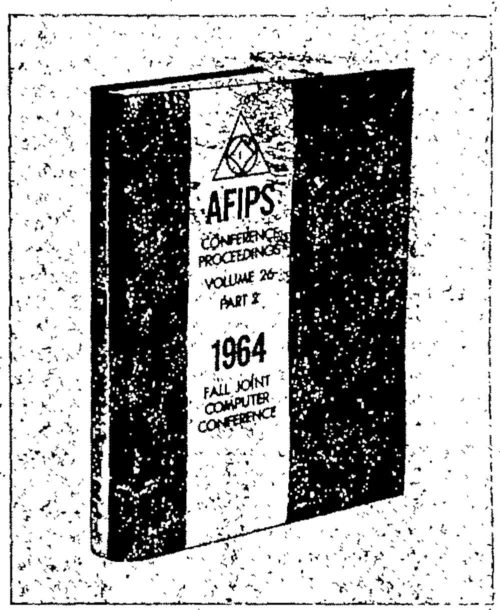
Return the coupon above. Payment with the order is requested.

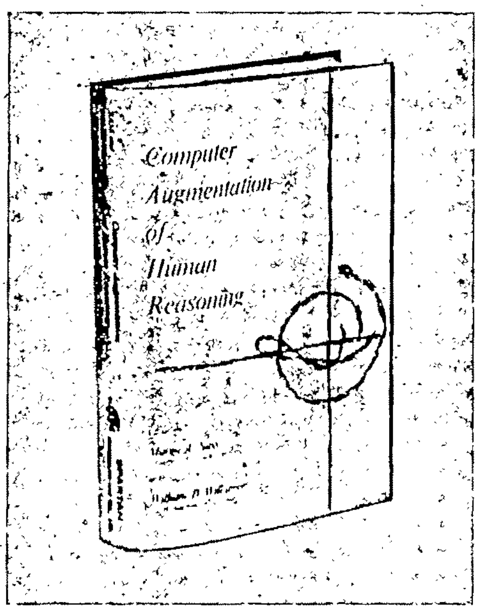
AFIPS 1964 FJCC, Part II of Vol. 26.

This second volume of papers presented at the 1964 FJCC features a comprehensive treatment of the high-speed computer systems presented by the various manufacturers. The papers include the new concepts, the hardware and software characteristics and some of the problems in reaching a final design for the new equipment. Price of the book was included in the registration fee, and the book will be mailed to all attendees of the conference.

160 pages 8½ x 11 illus. $4.75

50% discount to AFIPS affiliated members.





COMPUTER AUGMENTATION OF HUMAN REASONING

Edited by MARGO SASS, Office of Naval Research and W. D. WILKINSON, Bunker-Ramo Corporation.

Each advance in the computer sciences over the past twenty years has opened doors to new problems. The human problem solver traditionally has handled such matters by using past experience, trying alternatives, or applying intuition. Now, what is needed is a synthesis of the capabilities of man and machine. In this book experts exchange views on both the practical and the theoretical applications of the man/machine relationship.

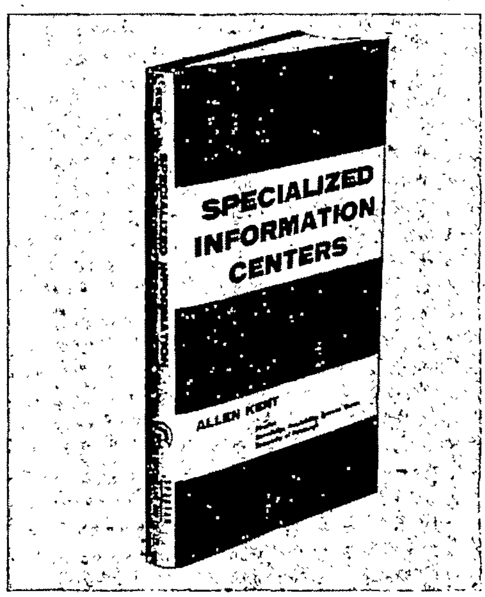
240 pages 6 x 9 illus. $5.00

SPECIALIZED INFORMATION CENTERS

By ALLEN KENT, Director Knowledge Availability Center University of Pittsburgh

This is the first book in the specialized information center field that provides detailed information on the centers, the unit operation, factors influencing their success or failure and comparative case histories of centers. For the computer specialist it is an especially useful book since the conceptual and software problems involved in the design and operation of the specialized information center are displayed.

296 pages 6 x 9 illus. $9.00



SPARTAN BOOKS, Inc. 1250 CONNECTICUT AVE., N.W. WASHINGTON, D. C. 20036

American Documentation

JULY 1965

Vol. 16, No. 3

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

SEMINARS ON SYSTEMS FOR THE INTELLECTUAL ORGANIZATION OF INFORMATION

A Conference on TECHNICAL INFORMATION CENTER ADMINISTRATION

THE SECOND ANNUAL NATIONAL COLLOQUIUM INFORMATION RETRIEVAL

ELECTRONIC INFORMATION HANDLING

CONFERENCE ON THE ORGANIZATION & UTILIZATION OF INFORMATION

SHORT COURSE INDUSTRIAL INFORMATION SPECIALISTS

A LECTURE SERIES IN INFORMATION SCIENCE 1964–1965

AMERICAN DOCUMENTATION

American Documentation is a publication of the American Documentation Institute. Its purpose is to be a scholarly journal in the various fields of documentation and to serve as a forum of discussion and experimentation.

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), one copy in the institutional membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

Instructions to authors are given in Volume 14, No. 2 (April 1963). All manuscripts, brief communications (1000 words or less), announcements, or letters should be sent to:

Mr. Arthur W. Elias, Editor
American Documentation
Institute for Scientific Information
325 Chestnut Street
Philadelphia, Pennsylvania 19106

Correspondence on reviews should be addressed to:
Dr. Theodore C. Hines, Review Editor
American Documentation
54 North Drive
East Brunswick, New Jersey

Correspondence on Literature Notes should be addressed to:
Mr. Charles P. Bourne, Literature Notes Editor
American Documentation
Stanford Research Institute
Menlo Park, California

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Science Abstracts,* and *Science Citation Index.*

American Documentation is entered for second class mailing at Washington, D. C.

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE

Vol. 16, No. 3 JULY 1965

Published quarterly by AMERICAN DOCUMENTATION INSTITUTE and printed by Spartan Books, Inc., 1250 Connecticut Avenue, N.W., Washington, D. C. Second class postage paid at Washington, D.C.

Editorial

"Conferences, Congresses, and Information"

In accordance with our custom, we have tried to link the cover illustration of *AD* with the theme of this editorial. The display of conference announcements shown is but a fraction of the material of this type received by your editor in the past year. We had believed that this proliferation is but a naturally occurring symptom to be expected in a young and rapidly developing field. We ascribed this to the need to communicate new developments more rapidly than is possible with other media, or a natural way of filling the educational and curriculum gaps noted in the ADI-Committee on the Organization of Information report found in this issue.

Quite possibly these beliefs are valid, for certainly these mechanisms have much to offer to neophytes, theorists, and practitioners alike. Still . . . an appraisal is certainly in order particularly on inspection of the remarkable similarities that can be detected in the subject matter, themes and, alas, even the speakers. Apart from geographic location, a curious sameness exists. When one may freely transpose institutional logos and come up with the same conference . . . in Pennsylvania Dutch, "It wonders me."

The prices of these sessions are a source of wonderment also. One need not look to Vietnam for examples of escalation, for a random sampling can indicate price variations in orders of magnitude.

Having just presided over one such conference, your editor is in a curious position with relation to this topic. Of course one's own efforts are always exceptions, but it is just this fact that started the train of thought put forth here. The sponsoring organizations, be they academic or commercial, have a heavy responsibility to conference registrants, the agencies and companies that foot the bills for the registrants, and most important, to themselves. For in the last analysis, irresponsibility, while it may wound innocent bystanders, almost certainly does the most damage to the perpetrator.

As with all trains of thought, this one is subject to derailment. The editor invites other engineers, firemen and switchmen to comment. He particularly would like to hear from some railroad management.

A. W. Elias

At its October 1964 meeting, the ADI Council unanimously endorsed a project to hold a symposium on education for information science. Some fifteen or more American universities now offer graduate training in that field. Conferences have been held on the subject at Georgia Institute of Technology, Western Reserve University, and the University of Pittsburgh. Each conference made its contribution, yet the fact remains that within the field of information science there is still little agreement on educational responsibilities to the profession and to all those fields depending on it.

The institutions conducting education for information science carry a heavy load of responsibility. But they should not have to bear this alone; they have a right to look for leadership to the U.S. professional society most directly concerned with documentation and information science. The American Documentation Institute should be the prime mover in its own field. Within its membership and affiliations is a broader spectrum of experience than can be found in any one institution. The planned symposium is the first attempt to squarely assume this responsibility—to hold a profession-wide meeting on what is perhaps its most important function—training and renewal of human resources in information science.

Much thought went into the form of the proposed symposium. A basic assumption made was that information science is a field of science in itself, but distinguished by the fact that it serves or supports communication in every other activity depending upon the stored message. Another assumption was that the principles and theory of this science have not yet been stated, and it is chiefly the lack of them which has prevented agreement on what information science is and, consequently, what to teach. Only when a field of knowledge has a well defined theoretical structure can we confidently lay out definite curricula. Until then the field remains more or less experimental. We cannot prescribe pre-requirements and courses until the fundamentals are known.

In order to provide some impetus in this direction and some focus for discussion, the approach taken was to write a "challenge paper." It was to be used as a point of departure, not to limit thinking.

To give the symposium some chance for careful discussions the number of attendees has been kept to about sixty. In order to counterbalance this limitation as far as possible, however, the ADI Council has provided that the challenge paper be printed in advance, that the Proceedings of the symposium be published as soon as possible, and that a final report be given in an open session (scheduled October 10, 1965) at the FID Congress.

Analogy Between Information Retrieval and Education

LAURENCE B. HEILPRIN
AND
FREDERICK L. GOODMAN

Staff Physicist, Council on Library Resources, Inc.
Washington, D. C.

Associate Prof. of Education, University of Michigan
Ann Arbor, Michigan

This work is presented as a challenge paper for a symposium on education for information science to be held by the American Documentation Institute in September, 1965. The paper suggests that both searching for information in a collection of stored messages and searching for information in the process of education have been subject to and shaped by one basic constraint—the very limited rates of flow of information into human sense channels. It is shown how information retrieval and education each have surmounted the same difficulty in the same way, by many-1 homomorphic transformations on messages which greatly reduce their word (or bit) content while preserving certain minimum invariants which identify the messages. From the homomorphic reductions, sensing time has been reduced further by means of the equivalence classes derived from the "vocabularies" in which the reduced messages are recoded. A quantitative partial model of information retrieval (to be presented in greater detail elsewhere) is suggested as homomorphic with education. Some possible applications to education in general and to education for information science in particular are discussed.

• Introduction

1. Both information storage and retrieval (ISR) and education (E) deal with the communication of ideas—messages. The messages which E conveys are of two kinds—informative or cognitive, and controlling or training. This analogy is mainly to the cognitive aspect: messages which convey information.

It will be obvious that, since we are dealing with communication, some of what follows might be restated in terms of "Information" theory, but we are not making this connection here or assuming knowledge in this area on the part of the reader.

2. All messages may be divided into two classes: those of short duration (SD) and those of long duration (LD). Of these, SD are the more basic in the sense that we could communicate with only SD messages but not with only LD messages. A message in its simplest form consists of two components. The first is some physical system, which we will call a "carrier," that is not in itself a message (examples: radio waves without the voice or music "intelligence"; a blank sheet of paper). The second consists in discriminable marks on the carrier such as images or sounds (SD) or printed letters or drawings (LD) which we will call "modulation."[1] The carrier can exist without modulation but not the reverse. In SD messages, the modulation varies in time. The "marks" on sound waves in direct speech change constantly at the ear—in fact they *must* die away (be attenuated) rapidly, if we are to be able to discriminate the next words or musical notes. If they persisted for even a short time more than they do, the sounds of successive speech or music symbols would become indistinct and blurred. Reverberation would destroy the meaning conveyed by the modulation. It is of the essence that SD messages be attenuated at least as fast as they pass into the sensor of a human recipient or of a machine-receiver. Somewhat more formally: the attenuation rate of the channel which conveys information to the sensor must equal or exceed the information rate of the sensor. By information rate is meant the time rate of change of fully discriminable "least units" of information such as word-symbols, or of their components, such as "bits."

Unless we refer specifically to bits/second or other rate units, by "sensing rate" we shall mean "words/second."

The reason why SD messages are more basic than LD is simply that when the message passes into the sensor of man or machine it must do so in SD form. Human sensory (afferent) and motor (efferent) messages travel by time-varying modulation to and from their destination or source—usually the brain. The same is true of machines which pass information through a sensor into some "decision" mechanism.

In contrast with that of SD messages, the modulation of LD messages persists for comparatively enormous time intervals. In order to achieve this extension into the time dimension, the carrier is restricted in most cases to a solid, and the modulation, instead of temporal, is spatial. Printed letters on a page store their contained message for long periods. They do so by extending spatially on the page. Naturally, since both SD and LD modulation exist in space and time, both are four-dimensional "marks." But the far shorter time duration of SD modulation makes us refer to it as "purely temporal," which it is not; and to LD as "purely spatial," which it is not.

Because of this constraint on communication—that a message must enter the sensor in the SD mode—all "stored" or LD messages must be convertible to SD form. This is indeed the case. It is also true that many SD messages are convertible to LD mode, but this was not always so—and the conversion is man's peculiar discovery. He found that information in LD mode can be stored, i.e., propagated into the time dimension, even beyond the life of the message sender.[2] The discovery enabled the cumulation of knowledge—the possibility for man's finite brain to tap a much large memory than his own.[3] Using this technique of storage as a tool, he erected science and civilization. But it is a basic constraint on use of the tool that stored messages be retransformed into the SD mode.

3. Since all messages are received in SD form a question of supreme importance is, what is the human rate of reception? And here lies, in our view, the major constraint on human communication: it is a very low rate—about 3 words per second or 40-50 bits/second, audio or visual, and somewhat higher for non-coded visual images.[1,4,5,6] Therefore, an important difference between SD communication (e.g., by speech, telephone, radio, television and sound-motion pictures) and LD communication (e.g., by books, articles, newspapers, reports, pictures, photographs) is that if SD messages (even when noisy) can be understood at all they are closely "matched" in rate to that of the recipient's input channel rate (V) and to his available access time for sensing (T). Were they not closely matched they would, per se, be unintelligible, and not receivable. By contrast, LD messages, because of their persistence in time, accumulate in "collections" of stored messages. Because the size of the collection soon greatly exceeds the amount of information which can be sensed at rate V in a reasonable access time T, the channel rates are hardly ever well matched. To state this arithmetically: if $N = nA$ where n is the number of main units (such as volumes, in a library) and A is the mean number of word-symbols per volume, then the access time to the collection

$$N/V = nA/V = T \tag{1}$$

is extremely long. For example, if n is one million volumes and A is one hundred thousand words per volume, while V is three words/second, then T would be (with continuous reading) about 1,000 years.

• Evolution of ISR and E a Consequence of Limited Information Sensing Rates

1. While it may seem fantastic to consider sensing an entire collection, as in the above example, this is what must be done if messages are to be located in a collection about which there is no prior knowledge. That is, there are two and only two ways to locate unknown messages in a larger collection, if we exclude sampling and require some kind of coverage of the entire collection:

(a) With no preparation—no prior sensing by anyone: sense the entire collection.
(b) With preparation—prior sensing by searcher or others:
  (1) search the entire collection;
  (2) "organize" the collection for information retrieval, and sense only a fraction of it.

Stated in this way, "organization" of a collection (which applies to both educational access to a body of stored information through a teacher's presentation or to information search in general) is simply a means of retaining coverage (which will be explained below) while reducing the available time T. If the sensing rate (V) remains constant, organization means arranging the collection so that it can be surveyed *in toto* but only some fraction N/K of its content need actually be scanned. How is this accomplished? Basically, there is only one way: by transforming the individual messages into much shorter ones so as to reduce, sometimes enormously, the amount of information (number of words, or of bits) contained. Examples: reviews, précis, digests, abstracts, catalog entries, index entries. All of these compress an original message into another having a smaller number of bits. The process is psychologically like what we do physically in order to view a nearby mountain as a whole. We move far enough away so that its entire outline appears in our field of view. But in order to accomplish

this, since our eyes admit only a finite amount of detail, we simultaneously lose nearby detail.

In recording the messages the "prior examiners" (e.g. abstracters, catalogers, indexers) observe one principle: they transform the message so that its identity is, in some sense, preserved. Message identity is kept by including in the shorter message certain components which (in their judgment) are essential: author's name, title, subject or subjects, etc. More technically, the reduced message is a *homomorphism* of the original. Let us suppose that two objects are compared. They appear similar but not completely so. To make them completely isomorphic (identical in function) one of the two objects (the more complex) is transformed by a many-1 transformation.* Since such a transformation is irreversible, some detail or "variety" in the more complex system is sacrificed. At a certain stage of simplification the transformed system becomes isomorphic with the second object. A homomorphism is a transformation involving a reduction in detail in one of two objects so that *after* the transformation the two become isomorphic, alike in all then-observable ways. Very generally, this is how our minds work in recognizing similarity and analogy. Given two things which appear to have the same relations between *some* components but not all, we *reduce* the internal relations (detail) of the more complex object so that it appears like the simpler object in *all* remaining respects. The number of similar relations is not increased, but the number of similar relations *relative to* the total relations discriminable in the more complex object is increased by the transformation, which erases non-similar details.[7]

Homomorphic reduction (HR) has two broad purposes:

(a) *Compression.* By substituting surveys, précis, informative abstracts[8] and the like for originals we in effect cover the same ground at increased speed, with no change in sensing rate.

(b) *Switching.* By using the symbolic entries in catalogs, indexes and vocabularies of keywords and descriptors, and the like, we manipulate the classes of which the reduced messages are members. In effect we divide the collection at any particular stage in the search into a part to be sensed and a part to be temporarily eliminated.

In education the need for the compressive type of HR if fully as great as, if not greater than, in ISR.

* A many-1 transformation or "mapping" of a set of objects S into a set of objects T is a correspondence that associates with each element s of S a single element t of T. That is, since nothing is said as to *how many* elements s of S are mapped into each element t of T, *many* elements of S can correspond to *one* element of T. After such a many-1 mapping, it is not possible, by examining only T, to reconstruct S by inversely mapping T on S. The original elements of S are no longer distinct. Therefore, a many-1 transformation is "irreversible" in the sense that it cannot be reperformed starting from T so as to recover all the detail (elements and their relations to each other) of S.

A lecturer or textbook author strives to substitute surveys, outlines, digests, examples and analogies for the "original" material. A switching system (a system which organizes pedagogical material) is used to permit the student to divide the total collection of material to be learned into a part to be sensed at a particular time and a part to be eliminated for the time being.

2. Before further comparison of ISR and E, let us complete sketching the steps in ISR alternative to sensing the entire collection. After the individual messages and perhaps some groups of messages have been homomorphically transformed, two physical operations are performed: the shorter collection is separated into an "image collection" or catalog; and the image collection of HR's is sensed. From it a vocabulary ("alphabet" of the HR messages) is formed. The members of the original collection are also tagged, not necessarily with this alphabet, but with some unique code (such as an accession number) associated with the corresponding HR message. The classes of which the reduced messages are members correspond to subsets of the collection of stored messages; hence, any operations carried out on the image collection effect like operations on the original collection, provided the tags preserve 1:1 identity. The actual search takes place in the image collection. Since the vocabulary of the image collection may be viewed as the "alphabet" from which the homomorphically-reduced messages are selected, each symbol in the vocabulary generates a class. Messages indexed by the same symbol ("indexed by" is here used as a general term meaning "containing the symbol in the abbreviated message") are members of "equivalence classes." The equivalence among members of the class indexed by the same symbol becomes better, the more precisely defined and standardized over all HR messages the symbols in the vocabulary.

The "alphabet" or vocabulary is now ready for search, in the case of coordinate search systems; or is further organized into compounds of logical inclusion and other relations, in more highly structured catalogs. Examples are catalogs based on partly ordered systems and hierarchies. No one rule or set of rules is used to organize the classes formed by the symbols in the "switching" HR system; what is best for one type of collection is not necessarily best for another. Thus the choice of indexing system and its logical manipulation will probably always remain a highly skilled professional selection.

The search is now performed by the searcher, who mentally constructs a "search prescription" or "search question," and matches it as closely as possible in the vocabulary of the catalog; or expands it in a thesaurus, a device for enlarging equivalence classes so as to insure wider access by forming the logical sum of two or more single-symbol equivalence classes. Corresponding to the search prescription in ISR is the teacher's or the student's

statement of a topic, theme, problem or hypothesis. The analog of the catalog and thesaurus is the set of pedagogical materials (such as texts and references) which the student and teacher use in the process of bringing relevant and pertinent information to the attention of the student. Each quest thus starts with a mental "prescription" of more or less limited extent; each prescription is modified as the search proceeds. Each body of messages available for search is, ultimately, the same collection—the stored (LD) knowledge of the world. The purpose of both the ISR and the E quest is not only to locate information but *to modify the search prescription.* In neither case need there be a high priority on the searcher's retaining located information, although it may be useful for some purposes. Similarly, in research, solving one problem usually redirects the search into a modification of the problem as originally expressed. Far from being merely a static starting point of entry to a search, a search prescription evolves dynamically as the search proceeds.

Returning to the actual logical operations performed on equivalence classes—there is an analogy between the range of methodologies in ISR and in E. In ISR search by coordination, the searcher himself combines the simple vocabulary codes into search prescriptions. By observing the number and quality of messages retrieved, he forms inductions (perceives relations) as to the proper combinations to use for finding desired subclasses of the collection. Analogously, the student using "progressive" or "inquiry" methods of education is relatively free to form his own search prescriptions, his own inductions, and conducts a search for knowledge with a minimum of constraint on the switching routines utilized. With more highly structured catalogs the ISR searcher is conducted by unseen catalogers and indexers to inductions selected by others as most probably useful. Likewise the student in a strict, preselected curriculum also has his inductions provided to a large degree. He "reaches the proper conclusions" set for him by the system. In each case the spectrum of constraint on search ranges from comparative freedom on the part of the searcher in choosing a logical path among the classes derived from the homomorphic reductions—to preselection of this path by someone other than the searcher.

The third and final stage in ISR is simply that of physically locating the desired subset of the original collection. This is easily done through the 1:1 correspondence between a code or tag on the image HR and a corresponding tag on the original. This code also corresponds to a physical "address" or location. The residue subset is then sensed at the channel rate V, usually in some time T well-matched to the available time of the searcher. When this "read-in" is completed, the entire collection has been scanned, albeit at the lower resolution of the switching collection of HR messages. In E this process involves moving from pedagogical material to original material and is often neglected at early stages of a person's education. Coping with "original sources" is the mark of the advanced student. The extent to which a student should be confronted with "original" information at an early age is a legitimate point of controversy among educators. The decision involves the value of having a student spend more time scanning original information and therefore of covering (spanning) less of the total collection than he could by scanning material which has undergone more intensive homomorphic reduction.

• Model of Information Retrieval-Educational Search

Suppose a body of knowledge is stored in a collection of n main units (e.g., volumes) which average A least units of information (e.g., words per volume). Let $N=nA$, and let V and T be the sensing rate and access time for sensing, respectively. If K is the average ratio of number of words per original unit (A) to the number in the homomorphic reduction (B), the $K=A/B$. Then

$$N=nA=KCVT \tag{2}$$

where C is the power of the classification system to divide the collection (e.g., $C=1$ corresponds to having no way to decide which part of it is to be sensed, so that all N must be sensed; while $C=10$ corresponds to a classification dividing power such that we can divide the collection into 10 parts and identify one part, $N/10$, as that to be sensed, the other $9N/10$ eliminated). Keeping $C=1$ (sensing the entire collection, at any value of K), we have:

Originally, at $K=1$ (no homomorphic reduction):

$$T=N/V \tag{3}$$

After HR to $K>1$:

$$T\ (K)=N/KV \tag{4}$$

Homomorphic reduction reduces the access time by dividing it by K. This example with $C=1$ is an oversimplification since we do not usually sense the entire collection at $K=1$ or at $K=K$. To sense the collection only in part (i.e., to use the equivalence classes which enable us to divide the collection into parts) we must pass from $C=1$ to $C=N/K$, a process much faster, in general, than sensing the entire collection, but which does (as suggested above) vary according to the "search path" or series of logical steps (sometimes called "search strategy"). The strategy will, of course, be restricted by the amount of built-in or fixed class connections in the class system.# In general the search time for sensing the

\# One principle underlies the use of greater or less pre-combination: it saves more time to manipulate classes with many members (i.e., with large class extensions) than classes with few members. Since class membership (extension) increases as class definition (intension) decreases, the use of single, non-compounded symbols (usually words) as the intension defining retrieval classes-in-extension tends to insure rapid logical elimination of significant fractions of the collection. In other words, the main reason for using classes is to use large classes. A catalog with a separate class for each volume would be no better than a book list. However, if the search objective is not only time saving but exhaustiveness, it may be more useful to use smaller classes (classes with smaller memberships produced by combinations of several symbols.)[1]

entire collection at any level K is the maximum of the various access times at that value of K. Therefore, we can conclude that it is usually possible to select a value of K such that the entire collection can be sensed in some practical time T. By accepting some compressed representation of the contents of the collection, the entire collection can be "spanned" by sensing.

Figure 1 shows a three dimensional model which represents the relations in equation (2). The axes of the rectangular tetrahedron are K, V, T. On the base plane ($K=1$) the diagonal lines correspond to values of N/C. The value for $C=1$ is the diagonal boundary of the base plane, the line $N/C=N$. As K increases, the outermost lines always correspond to $C=1$, so that lines in the outer bounding plane correspond to sensing the entire collection, at any value of K.

In the figure, an example of a double constraint is illustrated. $C=1$; and the sensing speed V is constant, $V=b$. This corresponds to the constancy of the channel rate of a machine or of a human. It is seen that on the base plane ($K=1$), the plane of original messages, the access time T is maximum. As K increases, T decreases. When K reaches the value $K=N/b$, then $T=1$. The entire collection can be scanned in one second, since the compression is so great that $N/(N/b)$ can be sensed at rate b in one second. Since the scale is logarithmic, no values lower than 1 are shown on the figure. However, the four apices of the tetrahedron are either theoretically unrealizable, or unrealizable at our present state of knowledge.

There are many features of this diagram which it is hoped to discuss elsewhere. For present purposes it is sufficient that the space within the tetrahedron can be used to represent any stage in a search path which involves sensing a body of literature. To represent a complete search path, i.e., a sequence of operations only some of which involve sensing the collection, requires another diagram, for example a plot of C against elapsed time.

It is apparent that the same variables apply to any instruction, by oneself or by another, in which the ability to sense symbols (i.e., read) is involved. Even wider application may be possible if V is extended to general rate of intake through a sense channel, in bits/second, not necessarily coded bits. In other words, the ISR model is based on variables that form an important subset of the larger set of variables involved in problem-solving. It is the common use of this subset in both areas that is the basis of the analogy between ISR and E.

• Concluding Remarks

A picture has been presented of ISR and of E in which, due to the accumulation of LD messages, any inquiry takes place among vastly more messages than can be sensed in the available time. Out of this constraint on use of stored information arose, we suggest, the main framework of the system of homomorphic transformations which underlie "bibliographic access" to stored literature. These compressions, as we have tried to show, reduce sensing time in two ways: (1) by permitting us to sense the homomorphic images in lieu of the originals, or (2) by permitting us to sense all of a collection (to scan-and-span) by means of the equivalence classes derived from the set of homomorphic reductions. In this view, most if not all of our presently-known bibliographic tools are simply alternative means of sensing an entire collection within a practicable time. The evolution of ISR systems is our response to one basic constraint on human sensing rates.

Both the structure of ISR and of E have been shaped by the same constraint. Each has been forced to compress messages—E in a manner possibly more varied than ISR—but each in response to the same basic limitation on channel flow of information into human sense channels. In each area, compression has enabled us to retain comprehensiveness in search of some intellectual domain—at the sacrifice of detailed knowledge.

The picture we have presented does not cover the specific way in which equivalence classes are or should be used. The logic of classes or its equivalent has been considered a prime factor in any theory of information retrieval and has often been considered a factor in pedagogical theory. Yet the present theory merely provides a framework for the application of *any* set of search strategies. We justify this in part with the observation that the ultimate result of *any* search by any logical system is the same—"division" of the collection of messages into smaller, identifiable parts which are to be sensed or rejected.

The introduction of machines which can sense symbols at rates higher than human is the beginning of a new era in information retrieval, and possibly in education. It suggests some until-now unheard of possibilities. At the present stage of information processing, we can sense between 10^4 and 10^5 bits/second, or, at about 30 bits per word, between 300 and 3,000 words/second. With such speed, it may be possible to exchange or "trade" compressibility K or divisibility C for sensing speed. This is shown clearly on the diagram of Fig. 1 (lines of constant N/CK). We may be able to remain on the plane $K=1$ of untransformed messages. Moreover, we may be able to remain on the diagonal line $C=1$, on which the machine reads totally "unstructured" collections. The relationship of high machine sensing rates to E is more difficult to specify. In E, the sensing channels must largely remain the human ones. An additional problem arises because the rate of formation of inductions limits the useful input sensing rate. That is, the rate of sensing (by humans) which we have been discussing until now is not a pure "sensory" rate. It is the slower of two rates, the human sensing rate by the sense organ, and the rate at which the afferent information can be centrally processed. Pierce and Karlin have

shown that the limit on human channel rates is not set by the sense organ, but rather by the internal "higher" processes which accompany the combined sensing and perceiving.[4]

One key to raising the human sensing rate is essentially to raise this internal limit. If we assume that such "higher" processing is largely a matter of forming inductions or perceiving relationships, then a role for machines does emerge. A basic barrier to "inductive" teaching has been that a teacher must respond to each individual's inductive efforts separately. The correctness of an inductive leap can be evaluated by tracing the consequences thereof deductively. Teachers can only indulge in this Socratic technique when the teacher-pupil ratio is very low. Because of a machine's large memory capacity and rapid sensing speed the deductive consequences of a wide variety of possible inductions can be programmed so that students can be informed quickly (and independently) of the implications of their inductions. This technique is already being exploited in educational experiments with simulations and advanced types of teaching machines. Machines which rapidly perform logical processes might be viewed as reducing sensing time by suggesting inductions. If we are able to perceive the inductions, the machines, in effect, permit us to sense larger collections in the same time.

To examine more specifically the "trade" between classification and scanning rate, we note that any improvement in an ISR system through utilization of a machine's rapid *scanning* (as distinct from its processing) capabilities can result in an improved E system. That is, a student may check his own inductive hypotheses without relying on either a teacher's or a classifier's preset deductive system. He can do this by ma-

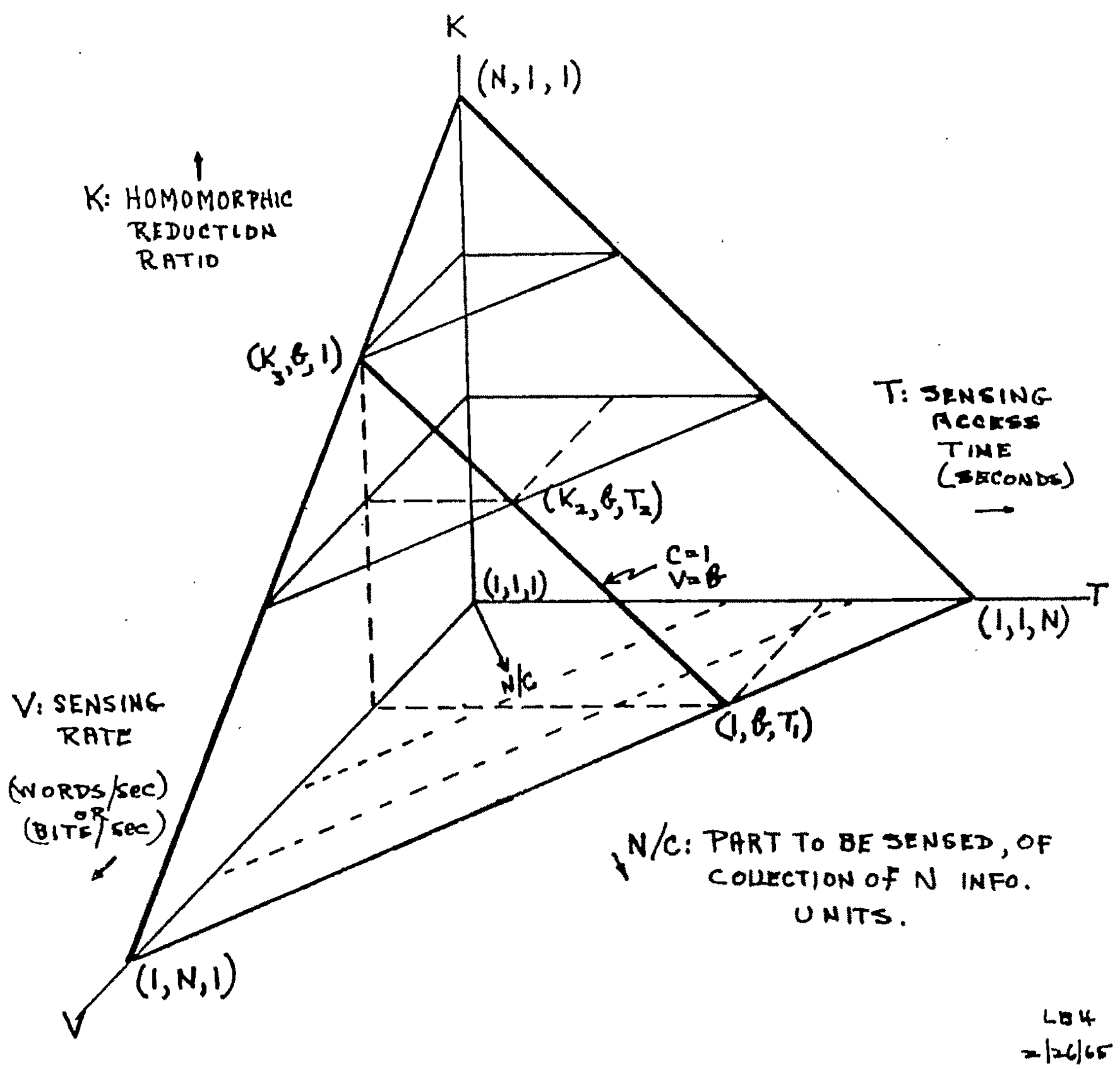


FIG. 1. ISR Model: N/C=KVT.
Example: (K,V,T) Values for Search under Constraints C=1, V=b, Log Scales.

chine-scan of *all* the information he is potentially concerned with rather than merely reading, in conventional fashion, the *parts* or sections declared most likely to be of interest by a predetermined switching technique. At an advanced level of teaching where creativity and independence are to be strongly encouraged, this "trade" may be an especially significant pedagogical aid.

Apart from emphasizing the possibilities of increased scanning speed by machines, it is possible to see other relationships between the V and T axes of Fig. 1 and systems of E. For example, scanning speed can be increased by "speed reading" techniques, in which the reading rate is doubled or tripled without loss of comprehension. The T variable, on the other hand, is traditionally manipulated by extending the length of formal education both by holding students in school longer and by encouraging specialization at an earlier age. If, however, ISR techniques are employed in E systems, a more efficient use of a student's total available access time may be realized not only by the methods suggested above but by reducing the time wasted on procuring information to be scanned. The ratio of time spent procuring reading matter to time spent reading it is probably very high. The ratio is probably even higher for teachers, once they leave academic communities. Individual pedagogical ISR systems devoted to improving the speed and independence of a student's access to information have been developed on a limited basis.[9]

So far we have suggested several applications of ISR theory and techniques to E in general. However, when we consider their application to education for ISR, namely, training information scientists, the full impact of the analogy emerges. Assuming that all bibliographic access arises through homomorphic transformations, and since all education involves them, information scientists and their teachers must be able to identify, create and organize homomorphic transformations. There is a clear challenge to develop instructional materials, methods and curricula based upon this unifying concept.

Such a unifying concept might also be related, perhaps embedded, in the Mathematical Theory of Communication of Shannon and school.[10] For example, the sensory bottlenecks could be re-stated as constraints on the matching of "external" to "internal" segments of the total communication channel. The constraints, in turn, might be expressed more generally in terms of actual as against potential entropy (relative entropy) and higher redundancy. Coding theory might shed light on homomorphic transformation. However, this transformation, central to the whole process, not only *recodes* messages to shorten them physically, but *alters* them, transforming them in a manner not predictable as yet by any theory. The shortened messages are often restated in words which are more abstract than those in the messages themselves. No provision for change in degree of abstraction has ever been quantitatively stated, and hardly qualitatively.[1] At best, we can say that the principle whereby the shortened messages are selected to represent the longer original ones is the invariance of *some of the meaning* of the messages, and their *relevance* to probable searches, in the judgment of the individuals making the transformations. Neither meaning nor relevance is explained by the present theory or by the Mathematical Theory of Communication.

In conclusion, we venture the generalizations that information retrieval may be very much a matter of teaching people how to search for information, while teaching people anything may be very much a matter of instructing them in the design and operation of information retrieval systems. If the first is reasonable, then those concerned with educating library or information scientists (or users of information systems) are of particular importance to their profession. Furthermore, they would profit by close attention to the homomorphic nature of searching and teaching processes. If the second is reasonable, then all educators could profit by close attention to the theory and technology of information retrieval.

References

1. L. B. Heilprin, "Information Storage and Retrieval as a Switching System," *Switching Theory in Space Technology*, Aiken and Main, Eds., Stanford University Press, 1963, pp. 298–332.
2. ——, "Toward a Definition of Information Science," *Proceedings, American Documentation Institute*, 1963, vol. 2, pp. 239–41.
3. ——, "On the Information Problem Ahead," *American Documentation*, Vol. 12, No. 1, 1961, pp. 6–14.
4. J. R. Pierce and J. E. Karlin, "Reading Rates and the Information Rate of a Human Channel," *Bell System Technical Journal* 36, pp. 497–516.
5. J. C. R. Licklider, K. N. Stevens and J. R. M. Hayes, "Studies in Speech, Hearing and Communication," Technical Report, Acoustics Laboratory, MIT, Cambridge, Mass., 1954.
6. H. Quastler and V. J. Wulff, "Human Performance in Information Transmission," Parts I and II, Control Systems Laboratory Report 62, University of Illinois Press, Urbana, Ill., 1955.
7. W. Ross Ashby, *An Introduction to Cybernetics*, John Wiley and Sons, New York, 1963, Chap. 6.
8. *Reports and Papers: The Royal Society Scientific Information Conference*. 21 June–2 July, 1948, London. The Royal Society, 1948.
9. F. L. Goodman, "Application of Information Retrieval Principles to Education," *Proceedings, American Documentation Institute*, 1963, vol. 1, pp. 51–52.
10. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Ill., 1949.

equations (2) and either (4) or (6) are used. The procedure used involves four basic steps and is outlined below. For the sake of brevity, S' in the following discussion refers to either S'_1 or S'_4, whichever is appropriate: S'_1 when $S \geq .5$, S'_4 when $S < .5$.

(a) S is converted to a point measured in radians, S', on the increment-decrement function. To accomplish this, S must first be converted to a value in the sine table, which is referred to as S_T in this discussion. Since S has a range of .5 for either of the sine curves used, and $0 \leq S_T| \leq 1$, S is converted to S_T by the equations $S_T = 2S - 1$ when $S \geq .5$ and $S_T = 1 - 2S$ when $S < .5$ Solving the equation $S' = SIN^{-1}(S_T)$ completes equations (1) and (2).

(b) The distance from S' to the nearest end-point of the function is calculated. The ranges of both sine curve quadrants cover 1.5708 radians. Therefore, this distance is always equal to $1.5708 - S'$.

(c) The distance calculated in (b) is divided by DIV and the quotient is applied (added or subtracted, depending on the type of response) to S' forming a new S'. From this it is clear that DIV is used in the increment-decrement function to modify the magnitude of the increases and decreases given to the profile word significance values. Its value should be adjusted to give as rapid an adjustment to the equilibrium values as possible and still have a stable feedback loop. By experimentation[9], a value of 8 seems to be a good compromise.

(d) The new S' is converted to S^*, the increased or decreased S. This is accomplished by solving equation (1) or (2), whichever is appropriate, in terms of S. In equation (1), $S = (SIN(S'_1) + 1)/2$ and in equation (2), $S = (1 - SIN(S'_4))/2$.

Step (a) is performed in equations (1) and (2). Steps (b), (c), and (d) are taken care of in equations (3), (4), (5), and (6).

The feedback process enables the profile word significance values to eventually reach an equilibrium point, which serves as a measurement of the word's real value to the profile. It may be found that a profile word can attain a higher equilibrium if another word is added to it to form a word cluster, thereby making the profile word a more specific description of the user's interests. The accuracy in describing the profile increases with the number of profile words that have reached an equilibrium. The more profile words with high equilibriums the better the results for the user. Our tests have shown that it takes anywhere from one to six feedback runs for a profile word to reach its equilibrium, depending on the frequency with which the word exists in the documents being disseminated.

All profile words and word clusters which matched document words, but for which a document was not selected for a notification, are increased similarly to a positive response but to a lesser degree. This is known as the slow increment. The introduction of the slow increment function into the SDI system is based on the

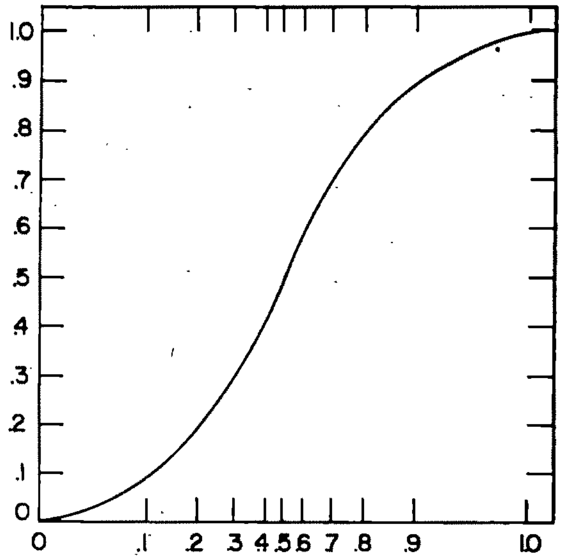


FIG. 4. Graph of Function Transpose.

supposition that profile words which consistently appear in documents must have some meaning to the user. A profile word occurring infrequently in the document set may be decreased to a significance value below the threshold because of unusual word composition in a document. If this word continues to appear infrequently in documents, the slow increment raises its significance value and allows the user to reconsider its significance through his responses. Conversely, profile words with significance values below the threshold which appear frequently in documents can be defined as "general" in relation to the interests of the user. The significance values of these words have reached an equilibrium at a low value and it is desired to have them remain there.

The slow increment given to a profile word significance value is computed as follows:

Let S = Profile word or word cluster significance value.

Let S'_4 = Profile word or word cluster significance value transpose on a fourth quadrant sine curve.

Let S^* = Profile word or word cluster significance value after the slow increment.

Let DIV = Variable divisor.

$$S'_4 = SIN^{-1}(1-2S) \tag{7}$$

$$S^* = (1 - SIN(S'_4 - (1.5708 - S'_4)/DIV))/2 \tag{8}$$

Equations (7) and (8) are similar to equations (2) and (4) explained earlier. This is because the slow increment given to a profile word significance value is computed using the increment-decrement function and can only be applied to positive responses and profile words with significant values $< .5$.

The only difference is in the value assigned to DIV. In the case of the slow increment function, DIV is set

equal to 16. The number of times the profile word appears in documents is added to the constant 16, thereby deriving the new *DIV* which is used in equation (8) to calculate the slow increment. This prevents the significance value of a high frequency word from oscillating wildly back and forth across the threshold level, which would otherwise be the case in a profile word with a potentially low equilibrium.

Figure 5 illustrates the value of the slow increment function. We have assumed that a profile word has been decreased twice in the first update run, so it has taken on a significance value of approximately .3 at the time of the second dissemination run. We have further assumed that this profile word continues to appear in the document sets used for succeeding dissemination runs. Line 1, Fig. 5, shows what would happen to the profile word significance value if there were no slow increment function. The word lies "buried" near the .3 significance value level. The only fluctuation in its significance value occurs when it appears in a document with another profile word such that the significance values of the profile words in combination were high enough to cause the document to be chosen. The slow increment given to the profile word if it was designated by Line 3, Fig. 5, is directly proportional to the number of times the word appears in the document set in each dissemination run. The significance value in this case oscillates back and forth across the .5 threshold level causing a wide variation in the number of documents chosen from run to run, and giving the profile word no chance to reach a stable equilibrium value. Line 2, Fig. 5, depicts the reaction of the profile word significance value with the Ames Laboratory SDI slow increment function. The significance value is gradually increased until it crosses the threshold level on the fifth update run. At this point it is free to attempt to establish its equilibrium value, dependent on the responses of the user.

We have also incorporated into our feedback function a cut-off which introduces time limits on the return of response cards. This cut-off is optional and all response cards which have not been returned and fall below this cut-off will be treated as negative responses.

A listing is kept of the selection numbers of all response cards returned where the "comment" "Port-a-Punch®" is punched.

(C) The user may request a listing of his profile or profiles at any time as shown in Fig. 6. All new profiles will be listed in the same way.

*Smutz, Morton Dr.
109 Link Bldg.
USER NUMBER — 00319

2. *Dissemination Function*

This section of the system consists of six programs which perform basically the following functions:

(a) Scan the document input and create single words.

The document input is acceptable only in the KWIC master record format as stated in "Keyword-In-Context (KWIC) Indexing General Information Manual," © 1962 by International Business Machines Corporation.

(b) Document words are matched with profile words, and the matched words are saved.

When matching profile words with document words, only the number of characters in the profile words is matched against the document words unless the user had specified further truncation or extension when he originated his profile. Through experimentation it was found that 16% of all matched words consisting of five or more characters were a result of truncated profile-document word match. Had these words not been considered matched, 33% of wanted documents would not have resulted in notifications. Obviously the number of unwanted documents would have dropped proportionately, however, we felt we would rather have the additional unwanted documents rather than lose some wanted documents. It should be noted that in the more precise profile words the percentage of truncated profile matches was much lower.

(c) After a matched word file has been created, the sum of the significance values of all words and word clusters within a given document unique to a particular user is obtained by using the following function:

$$0 < P_K < 1 \text{ [Profile Word or Word Cluster Significance Value]}$$

$$T_K = T_{K-1} + P_K - P_K T_{K-1} \text{ [Summation Function]}$$

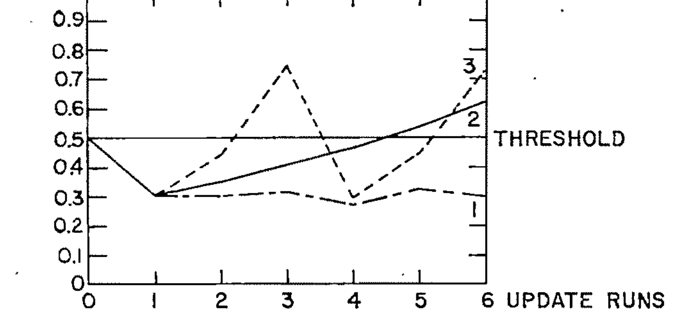


Fig. 5. Value of the Slow Increment Function.

Profile word	Char cnt	Profile No.	Sign. value
MONAZITE	08	00319 01	.5000
SOLVENT	07	00319 01	
EXTRACT—ON	10	00319 01	
RARE	15	00319 01	
EARTH	05	00319 01	.5000
RARE	15	00319 01	
EARTH	05	00319 01	
STABILITY	09	00319 01	
CONSTANT	08	00319 01	Neg.

Fig. 6. Ames Lab SDI User Profile Listing.

If we let T_K be the total probability that the user will want a particular document, it can be seen that the function above progressively calculates T_K by use of the formula for the probability of the union of two events.

(d) The summation of the significance values is compared with a threshold of .5.

If the summation is equal to or greater than .5, a notification is created. The threshold can be varied for experimental reasons. The notification (Fig. 3A) consists of a limited amount of the original document plus the threshold value, summation value, the first eight words or word clusters which were used in the summation, the user's name and address, response options, user's profile number, and prepunched selection number. The types of response options may be varied since the printing is controlled by one of the programs in this section of the system. The starting selection number for each run is fed in by means of a control card and must be kept unique and in ascending sequence. A corresponding internal selection number is also printed on the notification pending printer jams or correlation problems. These can be compensated for as already explained above.

(e) All matched words and word clusters are merged onto a matched word file.

The matched words and word clusters which appeared in documents which have become notifications will contain the selection numbers of their respective notifications.

(f) The author, title, source of each document, and profile user's name and number corresponding to each notification are listed.

This listing is used as a reference for securing documents if they are requested. It is contemplated, if the volume merits, to have the system create the actual order for the document from whichever source it is available.

3. *User Address Update*

This section of the system was an afterthought in the design of our SDI system, and its only function for the SDI system is to serve as a source of profile names and addresses.

(a) This section creates an employee directory with work and home addresses and phone numbers, job titles, supervisors, and employee numbers.

Options are available to update the directory for adding, deleting, or replacing single or multiple entries.

(b) From the employee directory, options are available to print a complete phone book, updated entries only, or mailing labels for particular employees.

• Ames Laboratory SDI Computer System Experiments

The document input[9] used in the experiments contained approximately one thousand documents. Included were research and development studies in the physical science fields, books on all phases of engineering, mathematics, and data processing, handbooks of tables, and even dictionaries. The title, abstract, and, in most cases, a few keywords were used in describing the documents.

The profiles used were oriented towards the interests of the people working for the Ames Laboratory. Each profile contained approximately twenty-five words or word clusters. The persons who selected the profile words had no prior knowledge of what was contained in the document input. About one third of the profiles used in the experiments were chosen by persons with experience in the subject area involved. The remainder were selected by persons not actually working in the particular subject area, but having a good concept of what people working in the area were interested in and were working on at the present time. So it was assumed that no significant change in the results occurred by allowing another person to simulate the profile of the actual profile user. The same persons who selected the profile words were responsible for determining, from notifications received, which documents were of interest.

The profiles were used as input, along with the document input, for the SDI dissemination run. Notifications were created from this run for those documents containing matched words which resulted in a total weighting factor greater than the .5 threshold as determined by the dissemination function. Responses were then made up for the notifications received and were used as input for the profile and matched word update run. The profile word significance values were increased or decreased according to the corresponding responses received. For the next iteration of this process the updated profiles were used as input for the dissemination run along with the same document input.

• *Experiment 1*

The purpose of this experiment was to measure the effectiveness of the SDI increment-decrement function.

Thirteen profiles were used covering thirteen different subject areas. A significance value of .5, the threshold level, was assigned to all profile words.

Logically, the feedback function would be most effective with words having significance values equal to the threshold level because of the immediate reflection of changes in word significance values on the dissemination of documents.

For an SDI system to be effective, two main conditions must be satisfied. The user must be interested in a high percentage of the documents he receives, and he must receive a high percentage of the desirable documents

	A	B	C	D	E	F	G	H	I	J	K	L	M
Run 1	38%	9%	50%	27%	19%	67%	80%	100%	67%	100%	100%	100%	67%
2	68	75	67	..	69	67	100	100	67	100	100	100	67
3	68	75	67	27	57	67	100	100	67	100	100	100	67
4	68	100	67	..	19	67	100	100	67	100	100	100	100
5	68	100	67	27	62	67	80	100	67	100	100	100	100
6	61	100	67	33	25	67	100	100	67	100	100	100	100

FIG. 7. Profiles—Percentage of Documents Wanted.

	A	B	C	D	E	F	G	H	H	J	K	L	M
Run 1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2	100	100	100	0	41	100	100	100	100	100	100	100	100
	100	100	100	44	100	100	100	100	100	100	100	100	100
	100	100	100	0	96	100	100	100	100	100	100	100	100
	100	100	100	0	48	100	100	100	100	100	100	100	100
	100	100	100	33	93	100	100	100	100	100	100	100	100
Total Available	20	3	2	3	27	2	4	1	2	1	1	1	2

FIG. 8. Profiles—Percentage of Wanted Documents

available to him. This experiment did not determine on an absolute basis the effectiveness of the system in selecting from the document input and directing to the appropriate user all documents of interest. It did, however, determine the relative effectiveness of the system under varying experimental conditions, as explained below. It was also possible to determine the degree of users' interest in the documents which they did receive.

Figure 7 shows the percentage of documents wanted of the total number of documents received for all thirteen user profiles. Most of the profiles reached a respectable percentage on the second iteration.

No notifications were created for Profile D on the second or fourth iteration because all the matched words from profile D were below the threshold level as a result of having been decremented. Notifications were created for profile D on the other iterations due to the fact that the "slow increment" raised the significance value of some of the words in profile D over the threshold level. The oscillations in the percentages for profile E were also a result of the "slow increment" raising the significance values of a few bad profile E words over the threshold level and causing some unwanted documents to be disseminated. It should be pointed out that the "slow increment" feature of the system was detrimental to the percentages of these two profiles because the document input remained constant in the experiment. The remaining eleven profiles retained their high percentages through all six iterations.

The percentages of wanted documents received of the total number of wanted documents available to the users from the disseminated output are shown in Fig. 8. The actual numbers of wanted documents available to the users are shown in the last line of Fig. 8. On all but two of the profiles, no wanted documents were lost over six iterations. The oscillations in the percentages for profiles D and E were the result of a few profile words being responsible for more unwanted documents than wanted documents and being decremented below the threshold level. On the next iteration many of these wanted documents were overlooked.

It is essential for a system to have a good balance between these two percentages if it is to be effective. After the sixth iteration, the system had reduced the volume of documents received by 40 percent and had lost only six percent of the wanted documents available.

From the results of this experiment, it appears that the increment-decrement function is quite effective. A marked improvement can be noted in the percentage of the documents wanted after the first iteration. The high percentages of documents wanted and wanted documents received are maintained throughout the experiment, and a satisfactory balance between these two percentages is achieved on each iteration.

• *Experiment 2*

The purpose of the experiment was to find how many iterations were necessary to reach an equilibrium, i.e. a leveling off of the results, when different significance values were used for the words in a profile.

Three of the profiles used in Experiment 1 were used for this experiment. Significance values of .1, .2, .3, .4, .5, .6, .7, .8, and .9 were assigned to the words in all three profiles, making a total of twenty-seven different profiles used for the experiment. By grouping the words with the same significance values together to form one profile, it was possible to determine whether or not one particular significance value reached an equilibrium sooner than the others.

The percentage of documents wanted of the total number of notifications received for all nine significance values of the three profiles is shown in Fig. 9.

Significance Value	.9	.8	.7	.6	.5	.4	.3	.2	.1
Run # 1	38%	38%	38%	38%	38%	89%	89%	100%	...
2	57	57	65	68	68	42	42	89	...
3	57	65	61	68	68	61	76	44	...
4	65	68	68	61	68	68	61	72	...
5	43	61	68	68	68	68	68	72	...
6	61	68	68	68	61	68	68	72	...
Significance Value	.9	.8	.7	.6	.5	.4	.3	.2	.1
Run # 1	50%	50%	50%	50%	50%	...	...	...	...
2	50	50	50	50	67	...	...	...	...
3	50	50	50	67	67	...	...	...	...
4	50	50	67	67	67	...	...	...	...
5	50	67	67	67	67	...	...	...	...
6	67	67	67	67	67	...	...	...	...
Significance Value	.9	.8	.7	.6	.5	.4	.3	.2	.1
Run # 1	9%	9%	9%	9%	9%	0%	0%	0%	...
2	9	9	38	60	75	0	0	0	...
3	33	60	38	75	75	0	0	0	...
4	38	75	38	75	100	0	...	...	...
5	100	100	100	100	100	...	...	...	...
6	100	100	100	100	100	...	...	...	...

FIG. 9. Percentage of Documents Wanted of Three Profiles.

In Profile A, a very active profile, both the .5 and .6 profiles reach their highest percentage of documents wanted, 68%, on the second iteration. This does not occur in the .5 profile until the sixth iteration, illustrating that the .5 profile is more stable than the .6 profile. The .7, .8, and .9 profiles all reach the same 68% as the .5 and .6 profiles, but not until later iterations. The .2, .3, and .4 profiles all start out with high percentages, but drop sharply after the second or third iterations, eventually leveling off at the 68% level. The .2 and .9 profiles reach the 68% equilibrium sometime after the sixth iteration. No notifications are received by the .1 profile.

Profile B is a very inactive profile in the experiment. Only three of the words in the profile match any document words, so not many notifications are created. As a result of this, no notifications are received by the .4, .3, .2, and .1 profiles. The .5 profile levels off at 67% on the second iteration; the .6, .7, .8, and .9 profiles level off on the fifth iteration. No wanted documents are received by the .2, .3, and .4 profiles, and no notifications are received at all by the .1 profile.

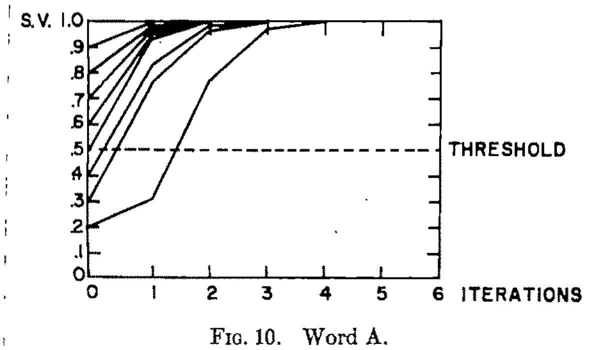


FIG. 10. Word A.

From the above three profiles, four specific words are selected to illustrate how an individual profile word reaches an equilibrium. Word A is a high frequency word, i.e. matches with document words quite regularly, that settles out to a high significance value. Word B is a high frequency word that settles low. Word C is a low frequency word that settles high, and Word D is a low frequency word that settles low.

Word A reaches its high equilibrium for the .5, .6, .7, .8, and .9 profiles on the second iteration, for the .3 and .4 profiles on the third iteration, and for the .2 profile on the fourth iteration. Since no notifications are received by the .1 profile, no change in the significance value of Word A is recorded at the .1 level.

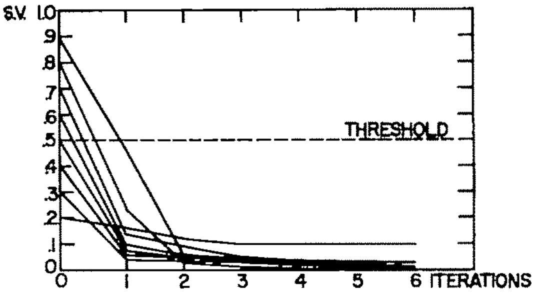


FIG. 11. Word B.

Word B, a very active profile word, drops to its low equilibrium on either the second or third iteration for all significance values. No notifications are received by the .1 profile.

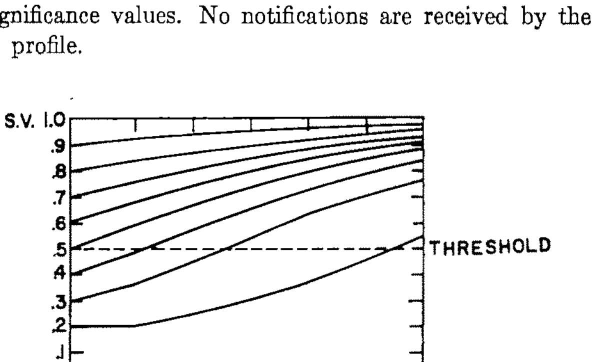


FIG. 12. Word C.

All the profiles reach their equilibrium sometime after the sixth iteration in the case of Word C, a low frequency word. No notifications are received by the .1 profile.

In the case of Word D, a low frequency profile word which settles to a low significance value equilibrium, the .5, .6, .7, .8, and .9 profiles drop to their equilibrium sometime after the sixth iteration. No notifications are received by the .1, .2, .3, and .4 profiles.

In this experiment, the results show that a profile composed of words with significance values of .5 reaches an equilibrium faster than any other significance value

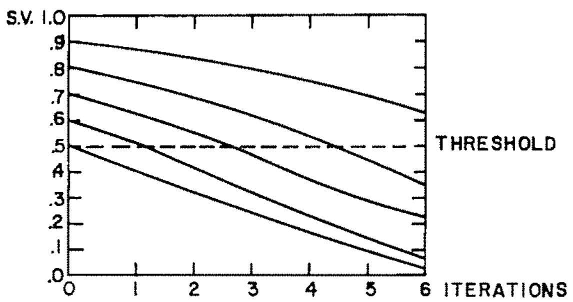


FIG. 13. Word D.

used, regardless of the volume of notifications received by the user. The .5 profile levels off sometime between the second and fourth iterations. The .6 profile is close behind the .5 profile in reaching an equilibrium.

The more active a profile word is in this experiment, the faster it reaches its equilibrium, regardless of whether its equilibrium is a high or low significance value. The high frequency words begin to level off sometime between the first and third iterations, depending on the initial significance value involved. The low frequency words require more than six iterations to reach an equilibrium.

• *Experiment 3*

The purpose of Experiment 3 was to determine the effectiveness of the "slow increment" given to matched words that do not appear in a notification. This occurred when the significance value of the matched words in a document, determined by the summation function, was less than the threshold level. Consequently, no notification was created. When this experiment was run, the value of *DIV* in the "slow increment" function (see Eq. (8)) was set to a constant 16, and the function acted independently on every matched word having a significance value below the threshold.

The three profiles used in Experiment 2 were also used in this experiment. However, this time only the significance value of .1 was assigned to the profile words, making a total of three profiles used in the experiment. By starting with a significance value of .1, it was possible to simulate what happens to a profile word that has been decremented to a low significance value.

Figure 14 shows the percentage of documents wanted of the notifications received for all three profiles used in the experiment.

	A	B	C
Run #1	....	...	....
2	38%	0%	....
3	72%	...	....
4	39%	...	25%
	69%	0%	43%
6	61%	...	43%

FIG. 14. Profiles—Percentage of Documents Wanted of Notifications Received.

Profile A required only one iteration to start receiving some notifications. Its percentage oscillated between the second and fourth iterations, but eventually leveled off at 68 percent sometime after the sixth iteration. Profile B, a less active profile than A, failed to receive any wanted documents during the six iterations shown in the table, but reached an equilibrium of 100 percent sometime after the sixth iteration. Profile C reached its highest possible percentage, 43 percent, on the fifth iteration, but began oscillating after the sixth iteration and continued to do so indefinitely.

Figures 15, 16, and 17 represent a closer look at Profiles A, B, and C. Each word in the profile that matched a document word is represented by a line on the graph of its particular profile. The numbers on the right-hand side of the graphs mark the significance values of the words after the sixth iteration. The corresponding numbers on the left-hand side of the graphs show the equilibrium values, if any, that are possible.

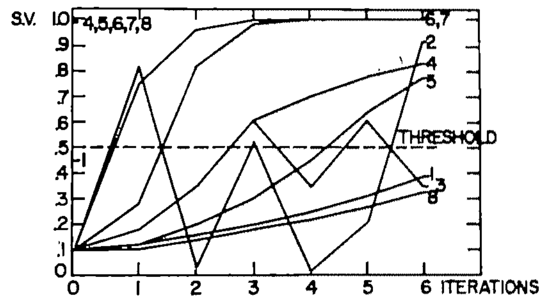


FIG. 15. Profile A.

Profile A contains eight words that matched with the document input. Words 6 and 7 reach their equilibrium on the third and fourth iterations, respectively. Words 4, 5, and 8 are not of as high a frequency as 6 and 7, so do not level off until sometime after the sixth iteration. The graph fails to show the entire picture of what happens to Word 1. It climbs slightly over the threshold level and then is decremented to a significance value of approximately .45. Here lies the value of the "slow increment." Word 1 lies buried and does not initiate any notifications, except when there is more than one matched word in the document. When the "slow increment" is applied to Word 1, its significance value is raised over the threshold level, so Word 1 can now initiate a notification. Words 2 and 3 have no equilibrium level, and oscillate above and below the threshold indefinitely. The graph is misleading in the case of these two words. It appears that both are buried and are being driven to a significance value above the threshold by the "slow increment." Actually, Words 1 and 2 appear in combination with enough other matched words to raise periodically their significance value over the threshold level without the aid of the "slow increment."

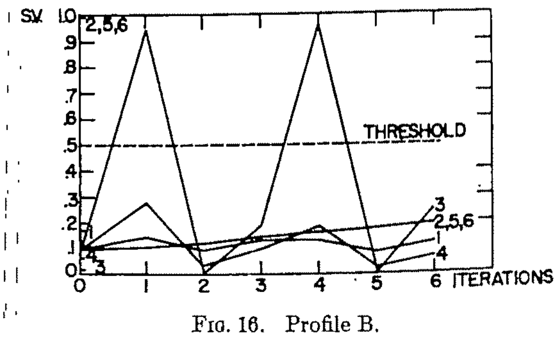


FIG. 16. Profile B.

In Profile B, Words 2, 5, and 6 reach their equilibrium sometime after the sixth iteration. Word 3 is a high frequency profile word that would have been buried on the second iteration were it not for the "slow increment." Words 1 and 4 appear not to be affected by the slow increment because they never reach the threshold level. Actually, these words are appearing in combination with other profile words to cause the dissemination of unwanted documents and are decremented to a low significance value. Then they are incremented by the "slow increment" to the point where they are able to cause the dissemination of these unwanted documents again.

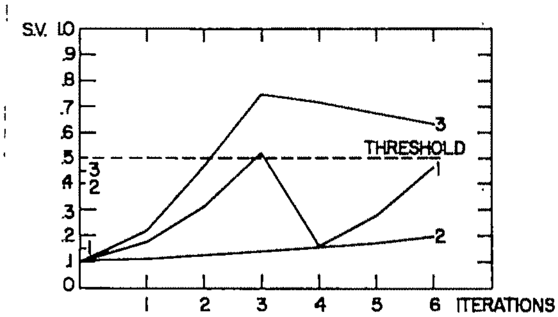


FIG. 17. Profile C.

All three words in Profile C show the effect of the "slow increment." The words oscillate above and below the threshold level, never being buried at their equilibria below the threshold because of the "slow increment." In the case of Word 2, this occurs sometime after the sixth weak iteration.

The value of the "slow increment" concept is easily seen from the results of Experiment 3. Without it, any profile word having a significance value below the threshold level is buried, except in a very active profile where the interaction with other profile words results in the creation of notifications. Figure 14 shows that some wanted documents are missed if profile words are buried below the threshold level. From the analysis of the results of this experiment, it was decided to change the "slow increment" function to make the amount of increase in the profile word significance value inversely proportional to the number of times the profile word appeared in the document input. In this way, profile words with significance values below the threshold which appear frequently in documents, and can be defined as "general" in relation to the interests of the user, remain at a low significance value. A profile word which continues to appear infrequently in the document set is raised by the "slow increment" function, thereby allowing the user to reconsider its significance through his responses. This change in the "slow increment" function is explained in detail earlier in this article.

• *Experiment 4*

This experiment attempted to find if there were any advantages of word clusters over single words.

This experiment was actually designed as two separate experiments. In the first part, twenty profile words were sent through a dissemination run. Then ten two-word clusters were formed from the original twenty profile words and matched against the same document input in a dissemination run. Then a comparison was made of the notifications received from the two runs. In the second part, the same operation was performed, first using a general word and then adding a relevant word to the general word to form a word cluster. Figure 18 shows the results of the first part of the experiment.

	Notif. received	Doc. wanted	% Doc. wanted
UNITERM	185	9	5%
2-WORD CLUSTER	26		31%

FIG. 18. Comparison of Notifications, Experiment 4.

The two-word clusters give a 26 percent better measure of accuracy on the percentage of documents wanted, while missing only one of the wanted documents available.

The remaining figures pertain to the second part of the experiment.

	Notif. received	Doc. wanted	% Doc. wanted
STATISTICS	21	14	67%
MATHEMATICAL STATISTICS	15	13	87%

FIG. 19. Results of Addition of Qualifier, Experiment 4.

Figure 19 shows the results of substituting MATHEMATICAL STATISTICS for uniterm, STATISTICS, in a profile from the subject area of statistics. The addition of the qualifier, MATHEMATICAL, improves the percentage of documents wanted by 20 percent while missing only one of the wanted documents available. By using the qualifier, unwanted documents are eliminated, e.g. publications from the Bureau of Labor Statistics, and the *World Almanac and Book of Facts,* which is a yearbook of statistics.

Figure 20 illustrates the effect of using DATA PROCESSING rather than DATA in a profile. The percentage of documents wanted increases by 51 percent, and only two wanted documents available are overlooked.

	Notif. received	Doc. wanted	% Doc. wanted
DATA	72	15	21%
DATA PROCESSING	18	13	72%

FIG. 20. Effect of Difference in Terms.

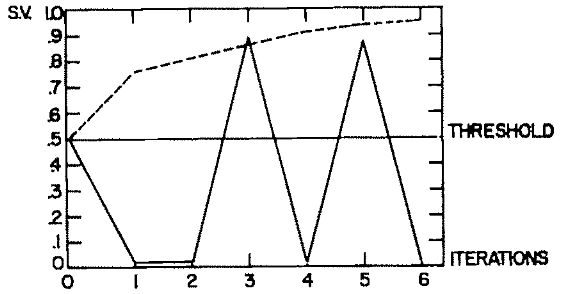


FIG. 21. Projected Significance Values.

Figure 21 is a graph of the projected significance values of the uniterm, DATA, and the word cluster, DATA PROCESSING, over six iterations. The word cluster is represented by the dotted line. The significance value of DATA PROCESSING reaches an equilibrium of approximately .99 sometime after the sixth iteration. The significance value of DATA oscillates above and below the threshold because of the "slow increment." It never reaches an equilibrium.

If the profile user wishes to include a general word in his profile, well and good. But the results of this experiment clearly point out that he can expect many unwanted documents if he uses too general a word for his purposes. When a qualifier is used to form a word cluster the improvement is evident from both the general case and the two specific examples shown in this experiment.

In summarizing our results to date, it can be said that the concept of feedback is proving to be an effective tool in the dissemination of information. The increment-decrement function appears to be successful in fulfilling one of the main aims of feedback, that of reducing the volume of notifications received without losing too many wanted documents. The "slow increment" function keeps a potentially good profile word from becoming extinct by periodically allowing for a check of its usefulness by the profile user.

• Ames Laboratory SDI Computer System Pilot Study[10,11]

Nine scientists were selected as users for our limited production tests based on their wide range of activities within the research program at the Ames Laboratory. Each profile varied in size and interest as shown below. Negation was not used in any of the profiles below.

Initially the newly derived profiles were run with 996 *Chemical Titles*[12] records as document input. The document records were approximately one third of the *Chemical Titles* bibliography tape used in the March 16, 1964, issue of *Chemical Titles* KWIC index. The tape records contained authors, titles, and sources of literature related to chemistry. Notifications of documents were returned to the users with a hard copy of the Chemical Title that was used as input. The users made modifications to their profiles and a second run was made with the revised profiles and the same document set. These tests were run to familiarize the users with the mechanics of the system.

The significance values of profile clusters, with a few exceptions, were assigned a value equal to our proposed threshold value of .5. The first production run disseminated 311 notifications of documents from a second set, same issue, of 1197 *Chemical Titles* records using the revised profiles. Figure 22 graphically describes the results. In the total process of disseminating these notifications only 8.8 percent of the total number of profile clusters participated. The participating profile clusters' significance values were subsequently increased or decreased when the responses were supplied to the profile cluster update. The percentage of participating profile clusters was small because the document input contained only author, title, and source records.

Table 1

Profile	Occupation	No. of one-word clusters	No. of two-word clusters	No. of three-word clusters	No. of four-word clusters
1	Chemist	19	254	5	0
2	Chemist	0	14	6	3
3A	Chemist	0	59	5	1
3B		0	5	3	1
4	Metallurgist	71	82	0	0
5	Physicist	3	101	3	0
6	Chem. Eng.	1	24	4	2
7	Physicist	0	7	5	0
8	Chemist	1	47	23	10
9A	Metallurgist	223	7	6	0
9B		210	78	13	1
9C		206	25	1	0
Totals		734	703	74	18
Average		61	50	6	2

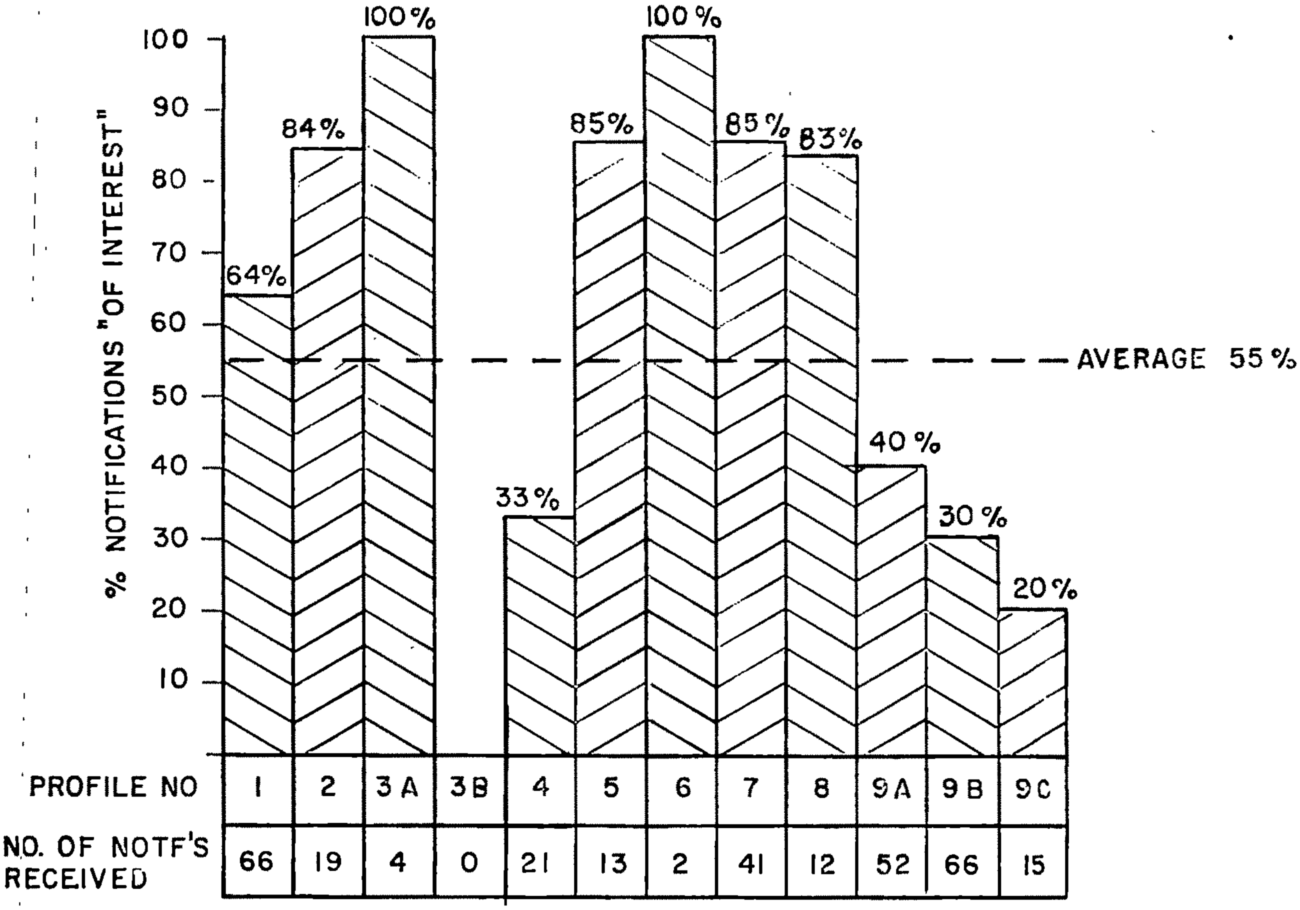


FIG. 22. Results of Run 1, 1197 Chemical Titles.

The second production run used a third set, same issue, of 1198 *Chemical Titles* records and disseminated 151 notifications of documents using the updated profiles from Run 1. The average "of interest" rose from 55 percent in Run 1 to 65 percent in Run 2 as shown in Fig. 23. The increase is a direct result of the fact that 42 percent of the profile words used in the selection of Run 2's notifications had been adjusted as a result of the feedback in Run 1. In the total process of disseminating these notifications, 12.8 percent of the total number of profile clusters participated. The running times of Run 1 and Run 2 and the feedback runs were 63 minutes and 42 minutes, respectively, on an IBM 7074, 10K, 12 tape drives machine.

The document set[18] used in the third production run contained 4176 documents. Included were research and development studies in the physical science fields, books on all phases of engineering, mathematics and data processing. Approximately 35 percent of the documents contained author, title, short abstract, and a few keyword records. The remaining documents contained author, title, and source records. The profiles used were the updated version of Run 2. The dissemination run produced 119 notifications of documents and the distribution is shown in Fig. 24. Of the total number of profile clusters, 8.8 percent participated in the selection of documents; 38 percent of these participating profile clusters had been adjusted as a result of Run 1 and Run 2. The average "of interest" dropped from 65 percent obtained in Run 2 to 59 percent.

A fourth production run was made using the updated profile of Run 3 and a third document set, *Nuclear Science Abstracts*[14], which covers the international literature on nuclear science and technology and is the most comprehensive abstract journal of this body of information. The abstracts were keyworded or indexed following the guidelines prescribed by the European Atomic Energy Community in the *Euratom-Thesaurus*. From 1861 documents, 659 notifications were disseminated to the users. Of the total number of profile words and clusters, 14 percent participated in the selection of documents; 21 percent of these participating profile words and clusters had been adjusted as a result of Runs 1, 2, and 3. The average "of interest" dropped from 59 to 52 percent as shown in Fig. 25. The percentage "of interest" described in these results is relative to the notifications of documents the users received. The users did not review all the documents used in this pilot study to determine the actual "of interest" documents from the total document set. The documents used in this

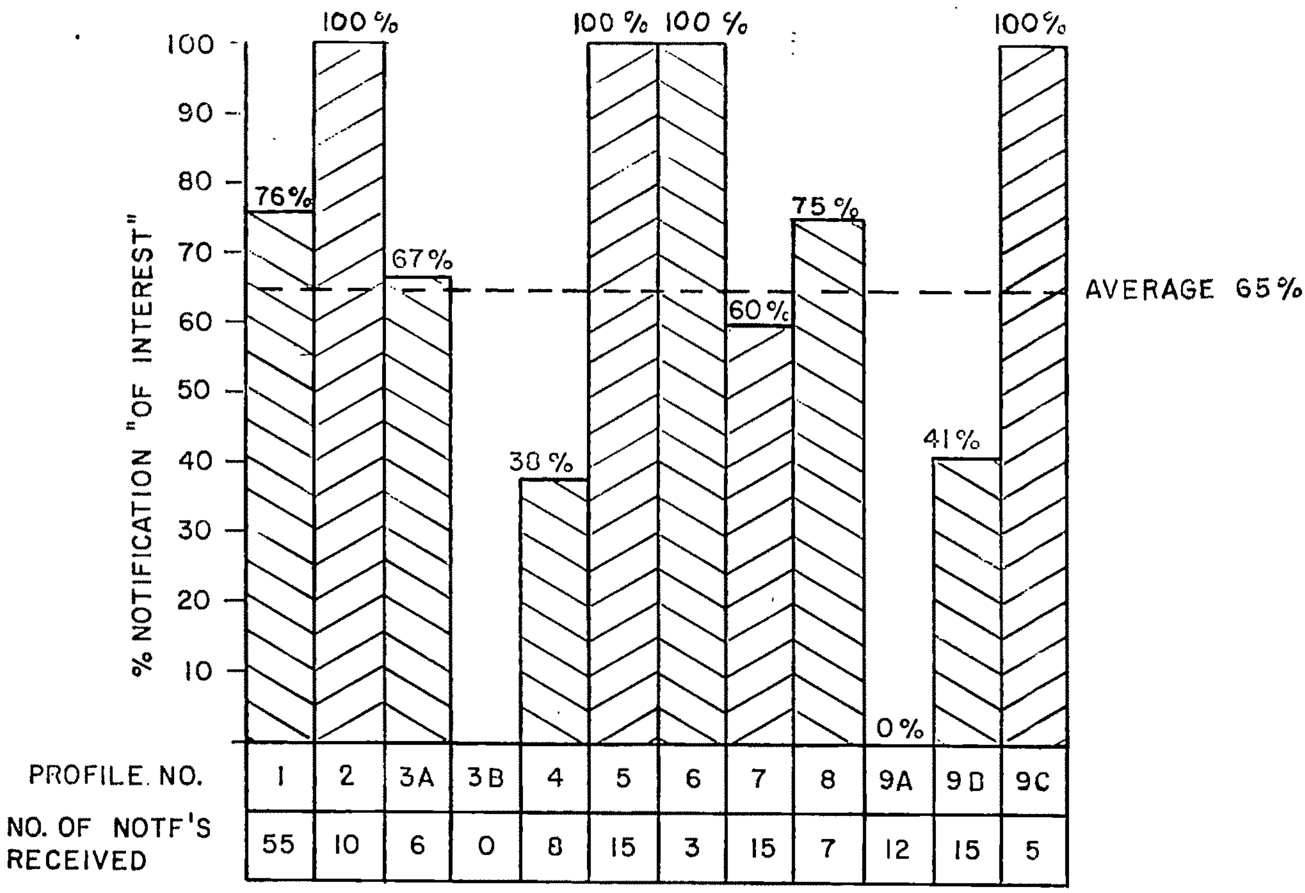


Fig. 23. Results of Run 2, 1198 Chemical Titles.

pilot study were not modified in any way except for format.

The operational results have indicated the system is sensitive to various types of document bases (e.g. author and title only, keywords, abstract, etc.) and will adapt to these to provide optimum performance. If the type of document bases changes from run to run the system is continually adapting and cannot produce optimum results. We have two theories as to what steps, taken independently or in combination, might correct this sensitivity: (1) combine the various document sets into average document sets to be disseminated for each run; (2) derive a mathematical model to analyze different document sets and "crank out" a threshold level appropriate to each unique document set.

The significance values of profile clusters which adapt to the more descriptive document texts will not be as sensitive to less descriptive document texts as a result of feedback. The performance of the system is not unlike a human in this respect; one might overlook a pertinent document because of inaccurate and incomplete title or abstract.

The scientists participating in this pilot project are generally pleased with the current performance of the system and have offered many helpful suggestions for improvement. A major point which pleases these scientists is the dynamics the system possesses in adapting to interest changes, provided the wide scope of interests was stated originally in their profiles.

The four production runs have involved approximately 244,110 document words and only 18 percent of the profile words participated in document selection. As this percentage increases with each production run, the statistics will improve.

The scientists in designing their profiles were responsible for synonyms. From the pilot studies it was found that the emphasis which seems to be placed on synonyms in information systems is unnecessary. It seems far better to make the user responsible for his word content because he is better qualified and can do a more efficient job than a computer program, no matter how sophisticated it might be.

This SDI system functioned successfully without negation and points out that negation is not vital to an SDI system. Negation has been incorporated into this SDI system to aid the user in designing a profile more simply.

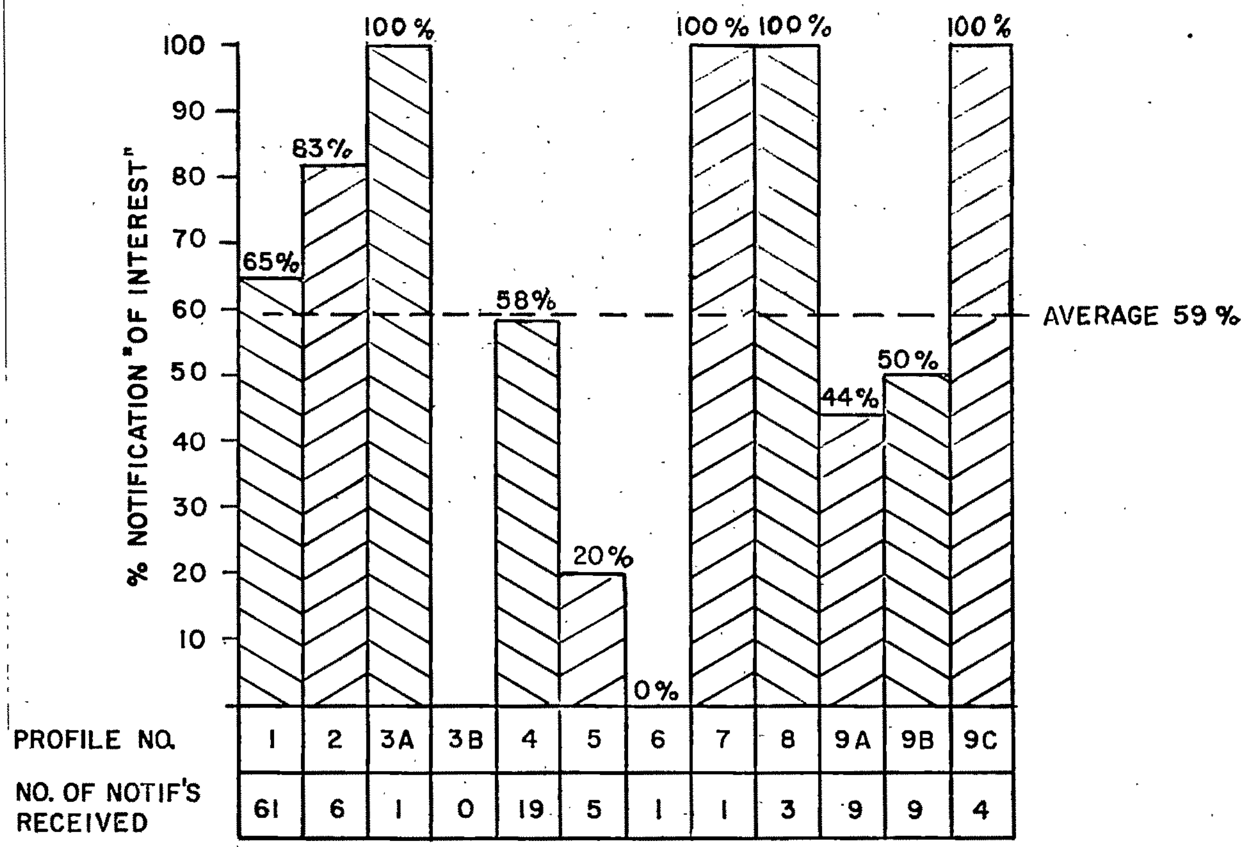


FIG. 24. Results of Run 3, 4176 Sandia Documents.

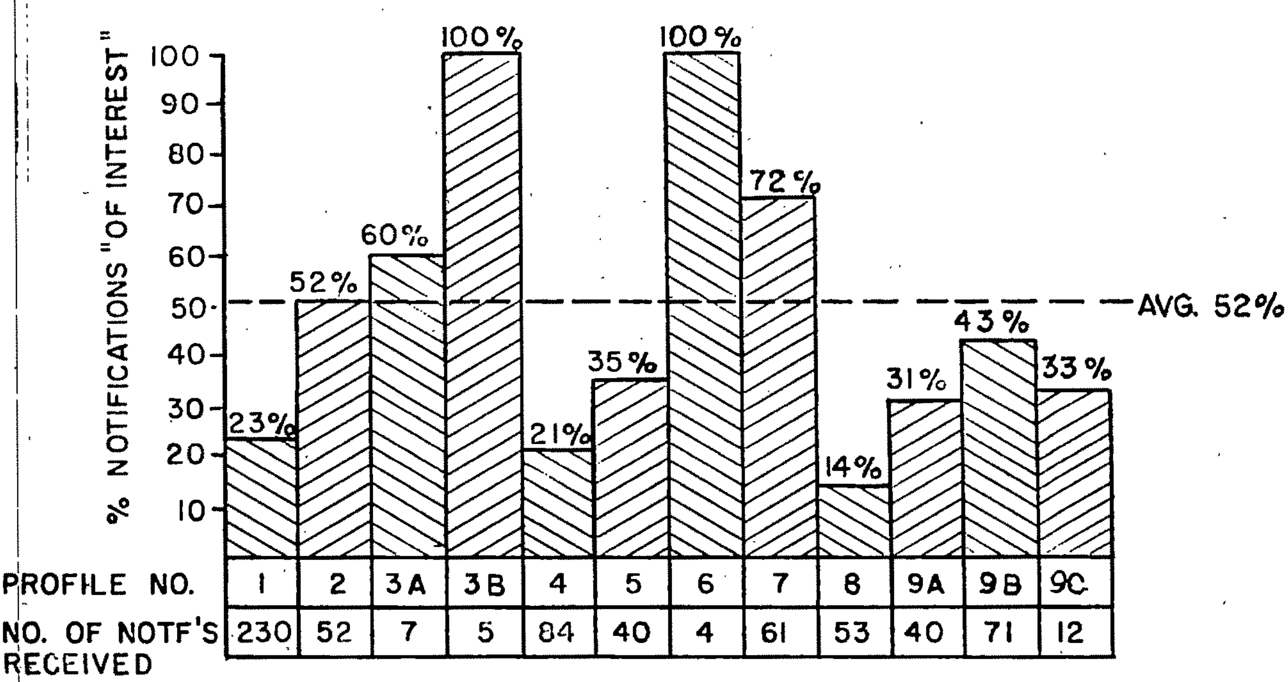


FIG. 25. Results of Run 4, 1861 "N.S.A. Keyworded Abstracts."

• Conclusion

From the results of the data obtained to date, it can be said that the concept of feedback is proving to be an effective tool in the dissemination of information. The increment-decrement function appears to be successful in fulfilling one of the main aims of feedback, that of reducing the volume of notifications received without losing too many wanted documents. The slow increment function keeps a potentially good profile word from becoming extinct by periodically allowing for a check of its usefulness by the profile user.

The system is designed with enough flexibility to accommodate the varying needs of the users without creating burdensome maintenance problems. Since the choice of words and word clusters is made by the user, the significance values which are assigned by the system provide the user with a valuable tool for the revision of his profile through additions, deletions and regrouping word clusters. This has been a very valuable by-product inasmuch as our users have all taken advantage of this to substantially improve their profiles. The various control options available to the user enable him to maintain sufficient influence in the management of his profile without impairing the effectiveness attributable to automatic updating of profile word significance values by the computer.

The benefits of the system are proportional to the amount of effort the user expends to fully utilize the potential the system has to offer him.

We feel that the regenerative process of feedback control loops is important and necessary to information processing. Systems of this type cannot be designed with a model which is "intuitively obvious." The semantics of the English language and the personality of the user are so variable that non-adaptive systems will usually be unstable or ineffective. We do not pretend we have the ultimate answer; however, we feel we have an effective model which can serve as a laboratory to study and improve our understanding of the complex problem of accurate dissemination, classification for storage, and retrieval of information.

In designing this SDI system we were seeking economy through the limitation of uncontrollable numbers of unwanted documents being disseminated to users; it is felt this has been partially accomplished. The cost of a notification varies between 5 and 12 cents; this, we feel, is not expensive and as hardware costs decrease and hardware capabilities increase the total operating cost of such a system will be reduced.

While the material benefits of this SDI system are difficult to measure, the scientists at Ames Laboratory who have participated in the system feel it will greatly increase their literature searching capabilities. It must be realized that this system does not replace existing methods of literature searching, but rather supplements existing methods of its users by making more information accessible.

This system is now in production with 68 profiles scanning approximately 6500 various document entries per week. These documents are purchased from Argonne National Laboratories and The Institute for Scientific Information.

References

1. C. R. Sage and D. R. Fitzwater, "Ames Laboratory Selective Dissemination of Information Computer System," U.S.A.E.C. Rept. IS-751, 1963.
2. International Business Machines Corporation, "General Information Manual, Keyword-In-Context (KWIC)," IBM Technical Publications Dept., 1962.
3. Institute for Scientific Information, Philadelphia, Pa. Current magnetic tapes (which were not used in this study) are employed by the Ames Laboratory for coverage of all U. S. Patents and the more than 1060 scientific journals as indexed in the *Science Citation Index—1965*.
4. H. P. Luhn, "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment," presented at the 126th Meeting of the American Association for the Advancement of Science, ACM, Chicago, Illinois, December 29, 1959.
5. S. A. Guyot, *Euratom-Thesaurus* (Keywords Used Within Euratom's Nuclear Energy Documentation Project), Brussels, February 1964.
6. A. B. Barnes, A. A. Briggs, J. Gauss, and A. Resmick, "SDI-5, An Advanced System for Selective Dissemination of Information," ACM, Proceedings of the 19th National Conference, 1964.
7. M. Taube, "Machines and Classifications in the Organization of Information," Documentation Inc., Technical Report No. 2, ASTIA AD No. 22426, Washington, D. C., 1953.
8. H. P. Edmundson and R. E. Wyllis, "Automatic Abstracting and Indexing Survey and Recommendations," *Comm. ACM No. 4,* May, 1961.
9. R. R. Anderson, D. R. Fitzwater, and C. R. Sage, "The Stability of Dynamic Feedback in an SDI System," Proceedings of the 19th International Guide Meeting, October 15, 1964.
10. C. R. Sage and D. R. Fitzwater, "Operational Results of an Adaptive SDI System," Proceedings of the 19th International Guide Meeting, October 15, 1964.
11. C. R. Sage, R. R. Anderson, and C. H. Cameron, "Ames Laboratory SDI Reference Manual," U.S.A.E.C. Rept. IS-940, 1964.
12. The Chemical Abstracts Service, *Chemical Titles,* The American Chemical Society, March 16, 1964.
13. Supplied by Sandia Corporation, Livermore, California.
14. U.S.A.E.C. Division of Technical Information, *Nuclear Science Abstracts,* vol. 18, no. 17, August 1964.

Testing, Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles

This paper evaluates the effect of system components such as links, roles, depth of indexing, vocabulary control, generics, and search strategy on recall, relevance, and cost of a patent index and compares retrievability and cost of coordinate indexing with classification.

BARBARA A. MONTAGUE

*Secretary's Department**
Information Systems Division
E. I. du Pont de Nemours & Co.
Wilmington, Delaware

• Introduction

A nonconventional technical-information system based on concept coordination using links and roles was designed and installed in the Plastics Department of the Du Pont Company in 1959.[1,2] After three years of operation, an interest developed in evaluating the system's performance.

The selection of the testing method was based upon the availability of two other systems, one nonconventional and one conventional, which contained documents common to those in the Plastics Department's files. Figure 1 illustrates the method used to evaluate system effectiveness.[3] The total file of an information system can be represented as displayed by the sum of X (Relevant References Retrieved) plus Y (Relevant References Not Retrieved) plus Z (Irrelevant References Retrieved) plus W (Irrelevant References Not Retrieved). Recall and relevance are two values discussed by Cleverdon[4] which can be used to measure an information system's performance. Recall is calculated in percent as one hundred times the quantity X divided by X plus Y, which is Relevant References Retrieved divided by the sum of Relevant Retrieved and Relevant Not Retrieved, or Total Relevant References. Recall is measured by addressing a series of questions to a system and examining retrieved references for relevance. The unknown quantity in this formula is Y, or Relevant References Not Retrieved. The method used to measure recall is based upon the assumption that, by addressing the same questions to two systems which contain common documents in their files and comparing references retrieved, most of the relevant references not retrieved by one system will be found by the other system and vice versa. In other words, the total relevant references to a given question will be the sum of the relevant retrieved between the two systems and 95% or more of the true number of relevant references will be found. Although the values calculated for recall for the two systems are relative, they lie within a range of 95% or better of the absolute values.

The second quantity which can be measured to determine system performance is relevance expressed in percent as one hundred times the quantity X over X plus Z, or the ratio of Relevant Retrieved to Total Retrieved.

A final word needed to describe the test method is a definition of relevance. A document was judged to be relevant if it contained information on the question for which it was retrieved. This is the same criterion used in the actual operation of the information system. No attempt was made to rank relevance quantitatively.

• Description of the Systems and the Tests

The three systems will be referred to as systems A, B, and C. Two tests were performed by comparing systems

* Formerly in Research & Development Division, Plastics Dept.

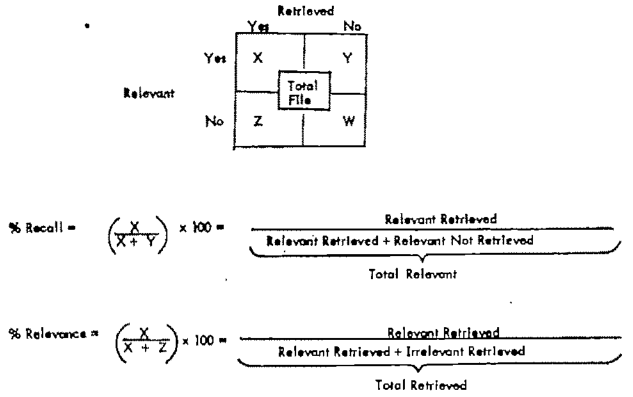


FIG. 1. Method For Evaluating System Effectiveness.

System A	System B
5000 Patents	811 Patents
Composition of Matter, Process, Articles of Manufacture, Apparatus	Composition of Matter, Chemical Process
Concept Coordination	Concept Coordination
Chemicals, Process, Properties, Equipment, Articles	Chemicals
90 Terms/Patent	70 Terms/Patent
Links	No Links
11Roles	5 Roles
C. A. Nomenclature	Patent Language
Generic Posting	No Generics
Claim vs. Disclosure	None
IBM 650	IBM 305 RAMAC
Intersections & Unions	Intersections & Unions

FIG. 2. Comparison of Systems A and B.

A and B, and systems A and C. Fig. 2 lists the components of systems A and B which were used for the first test.

Test No. 1: Systems A and B

System A contains 5,000 U. S. patents from the polyolefin art covering the period 1926-1960 and including composition of matter, process, articles of manufacture and apparatus patents. System B contains 811 patents from the same body of art covering composition of matter and chemical process only. The 433 patents common to the two systems served as the sample used in test No. 1. Both systems used concept coordination, with system A indexing the names of chemicals, processes, properties and equipment to an average depth of 90 terms per document for this sample of patents. System B indexed the names of chemicals to an average depth of 70 terms per patent, which means that system B indexed chemical terms to a greater depth than A, which distributed its 90 terms over a wider range of concepts. System A used links and a set of 11 roles, whereas system B used no links and five roles. System A employed vocabulary control by using *Chemical Abstracts'* system of nomenclature, provision for treatment of nonchemical terms for synonyms and near synonyms and generic posting of both chemical and nonchemical terms. System B, whose design was curtailed before completion, used no vocabulary control. The original design of the system had provisions for editing of the chemicals by fragmentation. This provided an opportunity to measure the effect of vocabulary control on recall and relevance. System A provided for differentiation between claim and disclosed information as an aid in performing domination searches. A used an IBM 650 computer with magnetic tape and random access for searching and B used an IBM 305 RAMAC computer. Search logic for both systems provided for performing intersections and unions.

System A	System B
3—Chemical Reactant	M—Monomer
7—Chemical Product	C—Catalyst
4—Special Agent	A—Aftertreatment Agent
5—Solvent, Medium	S—Solvent
6—Impurity, By-Product	I—Inert, Impurity
12—Subject of Patent	
1—Using	
2—Cause	
9—Effect, Object of Invention	
11—Physically Processed, Of	
00—Adjective	

FIG. 3. Role Systems.

Figure 3 compares the role systems used by A and B. System A uses eleven roles which differentiate chemical change from physical change using roles 7 and 11, and it provides for indexing of names of properties, processes, apparatus and articles using roles 0, 1, 2, 9, 11 and 12. B's role system provides for differentiation between monomer, catalyst, after-treatment, solvent and impurity and was adequate because only the names of chemicals were indexed.

	System A	System B
Indexing Time (Min./Patent)	83	60
Searching Time (Min./Question)		
Intellectual	20	66
Machine	24	4

FIG. 4. Comparison of Indexing and Searching Times.

Figure 4 compares the indexing and searching times of the two systems. System A's 83 minutes for input includes technical time required for indexing and editing. System B's 60 minutes represents technical time required for indexing only, as no editing was performed. System B required no clerical costs for indexing which consisted of simply underlining the terms on the patent itself and adding a role number. The indexed patent was then ready for keypunching.

The median time required for searching by A was less than that for B because of the fewer number of patents retrieved and the use of link-role abstracts versus patents for screening. System B, which used an IBM 305, required less time for searching than system A using a 650 because: (1) random access is faster than processing 6 magnetic tapes; (2) B had a smaller file of 50,000 postings compared to A's file of 2 million postings; and (3) the questions were formulated more simply using fewer terms by B than A.

The test consisted of drawing up 29 questions, both real and synthetic, and interrogating both systems. Real questions were obtained from searches made for the patent attorneys in our department and in another department of the company. Synthetic questions were drawn up based on the patents themselves. The patents retrieved by each system were screened for relevance by a staff of four literature analysts. The results are compared in Fig. 5. The total relevant was calculated to be the sum of the relevant patents retrieved between the two systems based, as mentioned earlier, on the assumption that most of the relevant patents will be found between the two systems.

	System A	System B
Total Retrieved	620	763
Relevant Retrieved	494	390
Irrelevant Retrieved	126	373
Relevant Not Retrieved	88	192
Recall = $\frac{\text{Relevant Retrieved}}{\text{Total Relevant*}}$	85%	67%
Relevance = $\frac{\text{Relevant Retrieved}}{\text{Total Retrieved}}$	80%	51%

*Total Relevant = sum of relevant retrieved between two systems

FIG. 5. Summary of Results of Test No. 1.

Recall was calculated for the two systems as a measure of the reliability or accuracy of the system. Likewise, values for relevance were calculated as a measure of false retrieval of the two systems. A statistical analysis of the data was made using a Wilcoxin assigned rank test which showed that there is no statistical difference between the two systems based on a comparison of the ratios of relevant retrieved to total relevant patents or, in other words, recall. System A was shown to be significantly better than system B based on a comparison of the ratios of Relevant Retrieved to Total Retrieved, that is, relevance.

Figure 6 presents a distribution of reasons why each system missed patents. System A missed patents due to intellectual indexing errors, not indexing deeply enough, searching too narrowly and clerical errors during processing. Because the people doing the analysis of test results did not have access to the input of patent indexing for system B, they could only deduce reasons for B missing patents. Based upon B's vocabulary listing and search strategy, which were both available, the lack of vocabulary

Cause	System A	System B
Indexing Error	5%	
Indexing Depth	3%	
Search Strategy	4%	
Processing Error	3%	
Vocabulary Control		5%
Generics		27%
Undetermined		1%
Total % Missed	15%	33%

FIG. 6. Reasons for Missed Patents.

control and generic search capability appeared to be the major causes for missed patents by system B. Some of the questions were generic in nature, and it is intellectually difficult to create a sum of all possible members of a generic class at the time of search.

Cause	System A	System B
Search Strategy	1%	6%
Generics	16%	2%
Processing Error	1%	1%
Links	1%	3%
Roles		9%
Undetermined		2%
Ref. to Prior Art		6%
Chemicals Only		20%
Vocabulary Control	1%	
Total Irrelevance	20%	49%

FIG. 7. Reasons for Irrelevant Retrieval.

In Fig. 7, a distribution of reasons for irrelevant retrieval is presented. The main factor responsible for irrelevance by system A occurred in one question for which a selective generic class was not available and the classes which had to be searched included concepts unrelated to the question. Half the irrelevant patents retrieved by system B were the result of indexing chemicals only, which resulted in an inability to distinguish monomers from polymers, homopolymerization from copolymerization, polymerization from chemical aftertreatment, or untreated polymers from treated polymers. Another reason for B's irrelevance was due to insufficient roles which did not provide differentiation between composition of matter, a product of a chemical reaction and a material which was physically processed.

Test No. 2: Systems A and C

System A was designed at the request of the Patents and Contracts Division of the Plastics Department to determine if concept coordination could solve the problem of time-consuming serial searching through their subject index to patents based on classification. Upon completion of the Coordinate Index, it was decided to compare the retrievability and time requirements of the two systems.

System C contains all 5,000 patents indexed by system A and consists of a subject index with one or two cross references on abstract cards. These items records are filed by subject in patent number order. To make a search one or more subjects are selected and each abstract is read to locate relevant patents. Searches are also made by system C using the patents themselves filed by subject.

Test No. 2 consisted of preparing 33 questions to be addressed to both systems. A random sample of sixteen of the chemical process and composition of matter questions used in Test No. 1 was selected and seventeen more questions were drawn up covering other patent types such as physical process, apparatus, and articles of manufacture. Some of these 17 questions were real searches generated by the Patents and Contracts Division and others were synthetically derived from the patents.

The 33 questions were run by system A using the 650 computer followed by screening for relevance by information scientists. The questions were run by system C by dividing them among 12 patent attorneys. The patents considered relevant by each system were compared question by question. A list of patents retrieved by system A but not by C was sent to the patent attorneys for final judgment concerning relevance and for reasons why relevant patents were missed by system C. A similar review was made for system A. Results of the test are shown in Fig. 8. Search times were calculated as medians rather than averages because of the skewed distribution of the data.

	System A	System C
Search Time (Hrs./Ques.)	1.1	5.1
Total Retrieved	2,533	522
Relevant Retrieved	1,165	388
Irrelevant Retrieved	1,368	134
Relevant Not Retrieved	92	869
Recall	93%	31%
Relevance	46%*	74%

*Chemical Questions 80% Relevance
Nonchemical Questions 20% Relevance

FIG. 8. Summary of Results of Test No. 2.

The high number of irrelevant patents retrieved by system A is due to the nonchemical questions on physical processes, apparatus and articles of manufacture, which cannot be characterized as rigorously as chemical questions. The relevance figure of 80% for chemical questions alone compares favorably with relevance data obtained in Test No. 1, which was limited to chemical questions.

An analysis of reasons for missing patents is illustrated in Fig. 9. Results of the test show that half of the 92 patents missed by system A were missed because of indexing errors or insufficient depth of indexing, and the other half because the search questions were framed too narrowly. Of course, these two causes are interrelated in a continuum, but it was possible in most cases to assign the reason for a missed reference to one or the other. System C missed most of its patents, the attorneys reported, because of input deficiencies which they ascribed to insufficient information on the abstract card to alert the searcher, no abstract card in the file, or no patent in the books. The reason assigned to searching was not searching enough subclasses or, in other words, searching too narrowly.

Cause	System A	System C
Indexing	45	703
Searching	47	166
Total Patents Missed	92	869

FIG. 9. Reasons for Missed Patents.

In addition to testing the relative performances of systems A and C, an economic evaluation was made. Figure 10 presents a comparison of input and search costs for the two systems. Input costs for system A include a little over one hour of technical time per patent for indexing and vocabulary editing, about half an hour of clerical time and 3 dollars per patent for keypunching, tabulating and computer updating costs. System C input costs are considerably lower, including about 5 minutes of abstracting and indexing time plus a half hour of clerical time. No machine costs are involved for C. Searching costs for system A totaled 32 dollars, as compared with 102 dollars for C. These data for system A reflect the philosophy of indexing by concept coordination which is a one-time, deep analysis of information at input which permits rapid, relatively inexpensive retrieval on demand. This can be compared with the classification index which involves inexpensive input but high-cost searching. The high cost of indexing by concept coordination can be justified by high-volume searching and more complete retrieval.

	System A	System C
Input per Patent		
Technical	\$9.50 @ \$8.10/hr.	\$1.30 @ \$10/hr.
Clerical	2.50 @ 6.00/hr.	2.70 @ 6/hr.
Machine (IBM 650)	3.10	—
Total	\$15.10	\$4.00
Searching per Question		
Technical	\$9.00 @ \$8.10/hr.	\$102.00 @ \$20/hr.
Machine (IBM 650)	23.00 @ 50/hr.	—
Total	\$32.00	\$102.00

FIG. 10. Cost Comparison.

A comparison of the results of Tests 1 and 2 is shown in Fig. 11. Test No. 1 compares the performance of coordinate indexing systems. The reasons for higher recall values for system A can be attributed to the use of vocabulary control, generic search capability, and deep indexing. The higher values which system A exhibited for relevance can be assigned to the use of links and roles to reduce false retrieval resulting from deep indexing.

Test No. 2 compares a coordinate index with a classification index, and results show that for the systems tested coordinate indexing is faster and retrieves more relevant information than the classification index. The test also shows that references to chemical information can be retrieved more selectively than references to nonchemical information.

	Test No. 1		Test No. 2	
	A	B	A	C
Recall	85%	67%	93%	31%
Relevance	80%	51%	46%*	74%

*Chemical Questions 80%
Nonchemical Questions 20%

FIG. 11. Comparison of Tests No. 1 and 2.

Effect of Links, Roles and Depth of Indexing on Recall, Relevance and Cost of Coordinate Indexing

The next phase of the testing program was designed to measure the effect of various components of a coordinate indexing system on its performance and cost. Deep indexing, links and roles were incorporated into the design of system A to increase recall of relevant information and decrease retrieval of irrelevant references. These tests were made to determine how effective these components are in obtaining the desired goal and what their utilization represents in terms of the economics of the installation and operation.

A series of tests shown in Fig. 12 was designed with the assistance of a statistician to measure the effect of three system components, links, roles and depth of indexing, on recall and relevance by comparison with the control. Test No. 3 measures the effect of links, Test No. 4 the effect of roles, Test No. 5 the effect of links and roles together, and Test No. 6 the effect of a limited selection of roles for searching. In each of the first four tests, depth of indexing was held constant at 40 terms per document. Test No. 7 measures the effect of various depths of indexing, holding the use of links and roles constant.

	Links	Roles	Indexing Depth
Control	Yes	Yes	Deep*
Test 3	No	Yes	Deep
Test 4	Yes	No	Deep
Test 5	No	No	Deep
Test 6	Yes	Limited	Deep
Test 7	Yes	Yes	Deep Medium Shallow

*40 terms per document.

FIG. 12. Evaluating the Performance of a Coordinate Index.

This series of tests was made on system A and the search questions employed were selected from the ones used in the comparison of system A with systems B and C discussed earlier. The total relevant patents for each question were those retrieved by all three systems with previous tests. Patents retrieved in answer to chemical questions on reactions and composition of matter were analyzed separately from those retrieved in answer to nonchemical questions involving articles of manufacture, apparatus, and physical after-treatment. The reason for this separate analysis was to compare the retrievability of chemical versus nonchemical information. In Test No. 2 relevance for nonchemical questions ran 20% as compared with 80% for chemical questions.

Test No. 3: No Links, Roles, Deep Indexing

Test No. 3 was run without links, with roles and deep indexing using a random sample of five chemical plus five nonchemical questions selected from previous tests and which retrieved a total of 500 patents. A random sample of questions was used instead of the total number to reduce the magnitude of effort involved in screening patents retrieved for relevance. The patent file was interrogated using the IBM 650 with a program especially written to suppress links during the search. The results are shown in Fig. 13. A control was run using links and roles, and one observes that the values for recall and relevance are somewhat lower than reported in previous tests. This is due to the fact that these data are obtained from a random sample of only ten out of the total questions used earlier. However, because this random sample of ten questions was used consistently throughout the link and role tests, the data can be compared with the control as shown. Two significant observations can be made from this comparison, first, that indexing without links had no significant effect on relevance or preventing false retrieval and, second, that the use of links did not affect recall by blocking retrieval of relevant patents. These observations hold for both chemical and nonchemical questions.

	Chemical Ques.		Nonchemical Ques.	
Test Conditions	Recall	Relevance	Recall	Relevance
Control— Links, Roles	83%	34%	83%	7%
Test No. 3— No Links, Roles	83%	33%	83%	7%

FIG. 13. Results of Test No. 3
No Links, Roles, Deep Indexing.

Test No. 4: Links, No Roles, Deep Indexing

Test No. 4 was run with links, without roles and with deep indexing using the same ten questions. The questions were run without roles by merging all roles of a term into a single set. The results are given in Fig. 14.

Results for the control and Test No. 3 are included in this figure for comparative purposes. Recall for chemical questions increased from 83% to 93% because running

the questions without roles resulted in retrieval of more relevant patients. Or, expressing it differently, indexing with roles blocked retrieval of relevant patents. No such reduction was observed with nonchemical questions. Roles blocked relevant retrieval of seven of 69 patents due to incorrect assignment of roles at the time of indexing. This fact points out that even using qualified personnel who understand the technology they are indexing, inaccurate assignment of roles occurs which affects recall of relevant references.

	Chemical Ques.		Nonchemical Ques.	
Test Conditions	Recall	Relevance	Recall	Relevance
Control—Links, Roles	83%	34%	83%	7%
Test No. 3—No Links, Roles	83%	33%	83%	7%
Test No. 4—Links, No Roles	93%	12%	83%	4%

FIG. 14. Results of Test No. 4
Links, No Roles, Deep Indexing,

The second observation to be made from this test is that indexing both chemical and nonchemical information without roles results in a significant reduction in relevance; that is, the use of roles does reduce false retrieval and costly screening time.

Test No. 5: No Links, No Roles, Deep Indexing

Test No. 5 was run using the same ten questions without links and without roles but with indexing depth held constant, and results are shown in Fig. 15. The data for Test No. 5 are similar to those from Test No. 4. There appears to be no interaction between links and roles and their effects are additive.

	Chemical Ques.		Nonchemical Ques.	
Test Conditions	Recall	Relevance	Recall	Relevance
Control—Links, Roles	83%	34%	83%	7%
Test No. 3—No Links, Roles	83%	33%	83%	7%
Test No. 4—Links, No Roles	93%	12%	83%	4%
Test No. 5—No Links, No Roles	93%	11%	83%	4%

FIG. 15. Results of Test No. 5
No Links, No Roles, Deep Indexing.

Test No. 6: Links, Limited Roles, Deep Indexing

In normal search strategy, a question is often written to search a term in more than one role to insure completeness of retrieval because some concepts can be indexed in more than one role. This suggests that roles may not be mutually exclusive. To test this concept, each of the ten questions was interpreted quite strictly and a single, best role was selected for each term searched. Links were used and depth of indexing held constant. The results are shown in Fig. 16. If one compares Test No. 6 with the control, it can be seen that for the chemical questions no appreciable effect is observed on recall or relevance using limited or strict application of roles. This indicates that the roles used to index chemical information are mutually exclusive and can be rigorously applied. The limited, strict application of roles for nonchemical information, however, results in a drastic reduction of recall with only a small increase in relevance compared with the control. These data indicate that redundant searching by merging several selected term-roles is required to handle the less precise, general, nonchemical information.

	Chemical Ques.		Nonchemical Ques.	
Test Conditions	Recall	Relevance	Recall	Relevance
Control—Links, Roles	83%	34%	83%	7%
Test No. 3—No Links, Roles	83%	33%	83%	7%
Test No. 4—Links, No Roles	93%	12%	83%	4%
Test No. 5—No Links, No Roles	93%	11%	83%	4%
Test No. 6—Links, Limited Roles	79%	36%	31%	9%

FIG. 16. Results of Test No. 6
Links, Limited Roles, Deep Indexing.

Test No. 7: Links, Roles, Varied Depth of Indexing

Exhaustive indexing with specific terms results in high input cost, large file size and, thus, increased machine costs. The logical argument supporting deep indexing is high retrievability. If one agrees that retrieving 93% of the relevant patents is an acceptable performance, the average depth of indexing of 40 terms per patent used for installation of system A is adequate. The objective of Test No. 7 was to measure the effect of indexing less deeply on recall of relevant patents. Three levels of indexing were selected: depth 1, which was the 40 terms per patent used in system A; depth 2, the terms used to index the claims of the patent plus the first half of the disclosure terms; and depth 3, the terms used to index the claims alone. A statistician selected a random sample of 90 patents from the total relevant patents retrieved by the two tests comparing system A with systems B and C. These patents were examined at each of the three depths to determine whether they would be retrieved in answer to the question which retrieved them at depth 1.

The results in Fig. 17 show that the values for recall at depth 1 are different from those cited for the previous link and role tests, as well as for the first two tests com-

	Chemical Ques. Relevant Ret'd.	Recall	Nonchemical Ques. Relevant Ret'd.	Recall
Depth 1	43	97%	34	83%
Depth 2	33	74%	28	68%
Depth 3	19	43%	18	44%

Depth 1 = claims + disclosure, or 60 terms patent/avg.
Depth 2 = claims + disclosure terms
Depth 3 = claim terms alone

Fig. 17. Results of Test No. 7
Effect of Depth of Indexing on Recall.

paring systems A, B, and C. This is due to the different sample of questions involved. The patents selected for this test sampled six of the chemical questions and ten of the nonchemical questions because the sample was based upon the total relevant patents and not on a distribution of questions. Recall for the six chemical questions sampled was 97% at depth 1 and successively lower at depths 2 and 3. A similar trend of reduction in recall with indexing depth is observed for nonchemical questions. These data are presented graphically in Fig. 18 with chemical and nonchemical questions plotted separately. Percent recall is plotted on the ordinate and depth of indexing in terms per patent is plotted along the abscissa at the three different depths. Extrapolation of the nonchemical line indicates that a depth of indexing of 35 to 40 terms per patent would be required to achieve 100% recall, whereas the depth for chemical questions is about right at 55. Another observation which can be made from this curve is that at the lower portion of the curve increasing depth has a definite effect upon recall but as you progress along the curve to deeper indexing levels, recall is less drastically affected. This kind of plot gives a quantitative measure of the depth of indexing required for an information system and shows that there is a point of diminishing return on the depth of indexing investment.

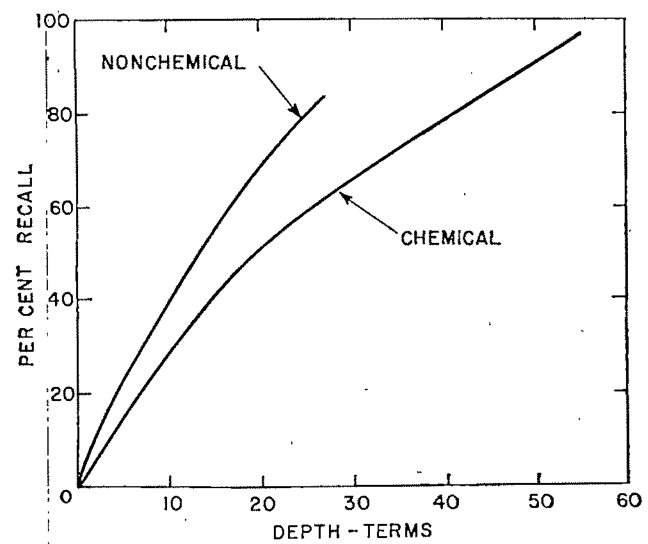


Fig. 18. Recall vs. Depth of Indexing.

Test No. 8: Cost of Links and Roles

A study of the incremental cost of indexing with links and roles was made by measuring the time required to add these two syntactical controls by five literature analysts who indexed 41 documents. The data are presented in Fig. 19 in terms of seconds per document required to add links and roles and the values are averaged for the five indexers. The average indexing time per patent using links and roles for system A was 36 minutes. Using the data shown in the figure, this means that 1.6 minutes were required to add links in indexing an average patent and 4.0 minutes to add roles. At an indexing rate of 8.10 dollars per hour, links cost 21 cents per patent and roles cost 54 cents per patent out of a total investment of 15 dollars. In other words, links represent 1.4% and role 3.6% of the total input investment.

Links (Sec./Doc.)	Roles (Sec./Doc.)	Terms (Sec./Doc.)	Total (Sec./Doc.)
44.7	114	859	1018
4.4% of Total	11.2% of Total		

System A Total Time 36 Min./Patent

Links required 1.6 min. at $8.10/hr. or 21¢/patent.

Roles required 4.0 min. at $8.10/hr. or 54¢/patent.

Total cost per patent $15.

Fig. 19. Time Requirements for Indexing with Links and Roles.

A study was also made of the effect of links and roles on the size of the file. The 5,000 patents were indexed with an average of 40 terms and 61 term-roles per document. This means that links and roles added 21 postings per patent, or increased the size of the file by 50%. To measure the postings added by links and roles separately, a sample of 200 patents was examined. Links added an average of 8 postings per documents, or increased the size of the file by 20%. Roles added an average of 13 postings, an increase of 32%.

Test No. 9: Cost of Depth of Indexing

The last test measured the variation of indexing depth with time. Five literature analysts indexed a total of 50 documents, noting the number of terms indexed at two-minute intervals. The results are shown in Fig. 20. Indexers A, B and D shown by the lower family of curves spend time initially analyzing the document before writing down indexing terms, whereas indexers C and E, the upper curves, begin indexing almost immediately. The curves show that 90% of the indexing is completed in 80% of the time invested and the remaining 20% of the time is not very productive in adding indexing terms.

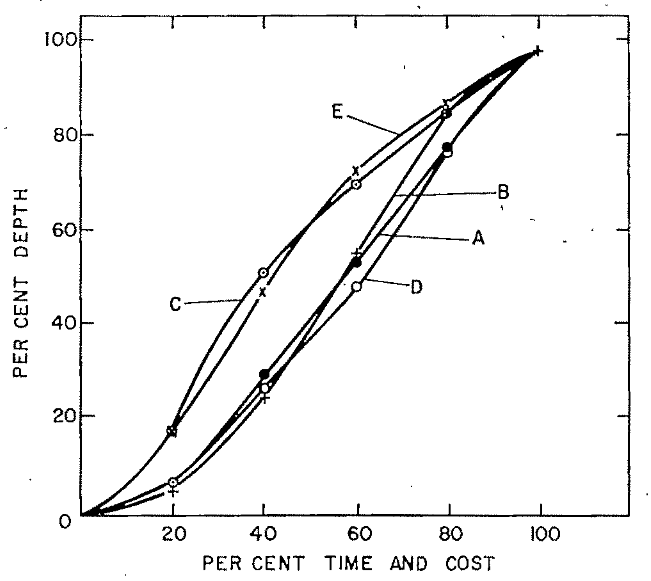


FIG. 20. Cost of Depth of Indexing.

• Summary and Conclusions

A series of tests was performed to compare the effectiveness of two coordinate indexing systems and a classification system. A second series of tests was made on one of the coordinate indexes to measure the effect of system components on performance and cost.

From the comparison of the two coordinate indexing systems it was concluded that a recall level of 90% is attainable by deep indexing, vocabulary control and provision for generic as well as specific searching. Links and roles improve relevance by reduction of false retrieval, and the use of these syntactical controls makes possible the attainment of a level of 80% relevance for retrieval of chemical information.

Secondly, the comparison of a coordinate index with a classification system shows that, for the two systems tested, coordinate indexing provides faster searching and retrieves more relevant references, and that the cost of coordinate indexing is higher at input and less for searching. Half the references missed by the coordinate index were the result of indexing errors and insufficient depth and the other half were due to searching to narrowly.

Third, the time required to index with links is 4% of the total time and links show no appreciable effect on recall or relevance of patent retrieval. These results are not surprising because patents are usually limited to one composition or process which can be indexed in one link.

Fourth, roles represent 11% of the total indexing cost and appreciably increase relevance. Errors in the application of roles result in reduced recall. The accurate use of roles requires technical indexers who understand the technology because indexing with links and roles is actually an abstracting procedure. Roles increase the relevance of retrieval of chemical information because they can be strictly applied. Roles need to be more rigorously defined and applied in order to reduce false retrieval of nonchemical information.

Fifth, depth of indexing appreciably affects both recall and cost, and data can be obtained to determine the depth necessary for a desired level of recall which can in turn be compared with a plot of depth versus time to measure the corresponding cost.

Sixth, chemical information can be retrieved with higher recall than nonchemical information because it is by nature more adaptable to rigorous description. For the same reason chemical information can be retrieved with a higher level of relevance, 80%, than nonchemical information, at 20%.

An evaluation of this types provides a great deal of insight into the operation of an information system. The results are useful in the optimization of existing systems and can serve as a valuable guide in the design of new systems.

References

1. J. C. COSTELLO, JR., "Storage and Retrieval of Chemical Research and Patent Information by Links and Roles in Du Pont," *American Documentation*, vol. 12, no. 2, 1961, pp. 111–120.
2. B. A. MONTAGUE, "Patent Indexing by Concept Coordination Using Links and Roles," *American Documentation*, vol. 13, no. 1, 1962, pp. 104–111.
3. J. KUHNS, "Word Correlation and Automatic Indexing," Ramo Woolridge Progress Report No. 2, C82–OU1, December 21, 1959.
4. CYRIL W. CLEVERDON, *Aslib Cranfield Research Project—Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems*, Cranfield College of Aeronautics, September 1960.

The Evaluation of Automatic Retrieval Procedures—Selected Test Results Using the SMART System*

The generation of effective methods for the evaluation of information retrieval systems and techniques is becoming increasingly important as more and more systems are designed and implemented. The present report deals with the evaluation of a variety of automatic indexing and retrieval procedures incorporated into the SMART automatic document retrieval system. The design of the SMART system is first briefly reviewed. The document file, search requests, and other parameters affecting the evaluation system are then examined in detail, and the measures used to assess the effectiveness of the retrieval performance are described. The main test results are given and tentative conclusions are reached concerning the design of fully automatic information systems.

GERARD SALTON

The Computation Laboratory of Harvard University
Cambridge, Massachusetts

• Introduction

The evaluation of information retrieval systems and of techniques for indexing, storing, searching and retrieving information has become of increasing importance in recent years. The interest in evaluation procedures stems from two main causes: first, more and more retrieval systems are being designed, thus raising an immediate question concerning performance and efficacy of these systems; and, second, evaluation methods are of interest in themselves, in that they lead to many complicated problems in test design and performance, and in the interpretation of test results.

The present study differs from other reports on systems evaluation in that it deals with the evaluation of automatic rather than conventional information retrieval. More specifically, it is desired to compare the effectiveness of a large variety of fully automatic procedures for information analysis (indexing) and retrieval. Since such an evaluation must of necessity take place in an experimental situation rather than in an operational environment, it becomes possible to eliminate from consideration such important system parameters as cost of retrieval, response time, influence of physical lay-out, personnel problems and so on, and to concentrate fully on the evaluation of *retrieval techniques*. Furthermore, a number of human problems which complicate matters in a conventional evaluation procedure, including, for example, the difficulties due to inconsistency among indexers or to the presence of search errors, need not be considered. Other problems, including those which have to do with the identification of information relevant to a given search request, and those concerning themselves with the interpretation of test results, must, of course, be faced in an automatic system just as in a conventional one.

The design of the SMART automatic document retrieval system is first briefly reviewed. The test environment is then described in detail, including in particular a description of the document file and of the search requests used. Parameters are introduced to measure the effectiveness of the retrieval performance; these parameters are similar to the standard recall and precision measures, but do not require that a distinction be made between retrieved and nonretrieved documents. The main test results are then given, and some tentative conclusions are reached concerning the design of fully automatic retrieval systems.

• The SMART Retrieval System

SMART is a fully automatic document retrieval system operating on the IBM 7094. Unlike other computer-based retrieval systems, the SMART system does

* This study was supported by the National Science Foundation under Grant GN-245.

not rely on manually assigned keywords or index terms for the identification of documents and search requests, nor does it use primarily the frequency of occurrence of certain words or phrases included in the texts of documents. Instead, an attempt is made to go beyond simple word-matching procedures by using a variety of intellectual aids in the form of synonym dictionaries, hierarchical arrangements of subject identifiers, statistical and syntactic phrase-generating methods and the like, in order to obtain the content identifications useful for the retrieval process.

Stored documents and search requests are then processed *without any prior manual analysis* by one of several hundred automatic content analysis methods, and those documents which most nearly match a given search request are extracted from the document file in answer to the request. The system may be controlled by the user in that a search request can be processed first in a standard mode; the user can then analyze the output obtained and, depending on his further requirements, order a reprocessing of the request under new conditions. The new output can again be examined and the process iterated until the right kind and amount of information are retrieved.

SMART is thus designed to correct many of the shortcomings of presently available automatic retrieval systems, and it may serve as a reasonable prototype for fully automatic document retrieval. The following facilities incorporated into the SMART system for purposes of document analysis may be of principal interest*:

(a) a system for separating English words into *stems* and *affixes* (the so-called "null thesaurus" method) which can be used to construct document identifications consisting of the word stems contained in the documents;

(b) a synonym dictionary, or *thesaurus*, which can be used to recognize synonyms by replacing each word stem by one or more "concept" numbers (the thesaurus is a manually constructed dictionary including about 600 concepts in the computer literature, corresponding to about 3000 English word stems); these concept numbers can serve as content identifiers instead of the original word stems;

(c) a *hierarchical arrangement* of the concepts included in the thesaurus which makes it possible, given any concept number, to find its "parent" in the hierarchy, its "sons," its "brothers," and any of a set of possible cross-references; the hierarchy can be used to obtain more general content identifiers than the ones originally given by going "up" in the hierarchy, more specific ones by going "down" in the structure, and a set of related ones by picking up brothers and cross-references;

(d) *statistical procedures* to compute similarity coefficients based on co-occurrences of concepts within the sentences of a given document, or within the documents of a given collection; association factors between documents can also be determined, as can clusters (rather than only pairs) of related documents, or related concepts; the related concepts, determined by statistical association, can then be added to the originally available concepts to identify the various documents;

(e) *syntactic analysis* and matching methods which make it possible to compare the syntactically analyzed sentences of documents and search requests with a pre-coded dictionary of "criterion" phrases in such a way that the same concept number is assigned to a large number of semantically equivalent, but syntactically quite different constructions (e.g. "information retrieval," "the retrieval of information," "the retrieval of documents," "text processing," and so on);

(f) *statistical phrase* matching methods which operate like the preceding syntactic phrase procedures, that is, by using a preconstructed dictionary to identify phrases used as content identifiers; however, no syntactic analysis is performed in this case, and phrases are defined as equivalent if the concept numbers of all components match, regardless of the syntactic relationships between components;

(g) a *dictionary updating* system, designed to revise the five principal dictionaries included in the system (stem thesaurus, suffix dictionary, concept hierarchy, statistical phrases, and syntactic "criterion" phrases).

The operations of the system are built around a supervisory system which decodes the input instructions and arranges the processing sequence in accordance with the instructions received. At the present time, about 35 different processing options are available, in addition to a number of variable parameter settings. The latter are used to specify the type of correlation function which measures the similarity between documents and search requests, the cut-off value which determines the number of documents to be extracted as answers to search requests, and the thesaurus size.

The SMART systems organization makes it possible to evaluate the effectiveness of the various processing methods by comparing the outputs obtained from a variety of different runs. This is achieved by processing the same search requests against the same document collection several times, and making judicious changes in the analysis procedures between runs. It is this use of the SMART system, as an evaluation tool, which is of particular interest in the present context, and is therefore treated in more detail in the remaining parts of the present report.

* More detailed descriptions of the systems organization are included in Refs. 1 and 2. Programming aspects and complete flowcharts are presented in Ref. 3.

Characteristic	Comment	Count
Number of documents in collection.	Document abstracts in the computer field.	405
Number of search requests (a) specific (b) general.	 0 – 9 relevant documents 10 – 30 relevant documents	 10 7
User population (requester also makes relevance judgments).	Technical people and students	about 10
Number of indexing and search programs used.	All search and indexing operations are automatic.	15
Number of index terms per document.	Varies greatly depending on indexing procedure and document.	(average) 35
Number of relevant documents per request (a) specific (b) general.		 (average) 5 (average) 15
Number of retrieved documents per request.	No cut-off is used to separate retrieved from nonretrieved.	405

FIG. 1. Test Environment.

The Test Environment

The parameters which control the testing procedures about to be described are summarized in Fig. 1. The data collection used consists of a set of 405 *abstracts** of documents in the computer literature published during 1959 in the *IRE Transactions on Electronic Computers*. The results reported are based on the processing of about 20 search requests, each of which is analyzed by approximately 15 different indexing procedures. The search requests are somewhat arbitrarily separated into two groups, called respectively "general" and "specific" requests, depending on whether the number of documents believed to be relevant to each request is equal to at least ten (for the general requests) or is less than ten (for the specific ones). Results are reported separately for each of these two request groups; cumulative results are also reported for the complete set of requests.

The user population responsible for the search requests consists of about ten technical people with background in the computer field. Requests are formulated without study of the document collection, and no document already included in the collection is normally used as a source for any given search request. On the other hand, in view of the experimental nature of the system it cannot be stated unequivocally that an actual user need in fact exists which requires fulfilment.

An excerpt from the document collection, as it is originally introduced into computer storage, is reproduced in Fig. 2. It may be noted that the full abstracts are stored together with the bibliographic citations. A typical search request, dealing with the numerical solution of differential equations, is shown at the top of Fig. 3. Any search request expressed in English words is acceptable, and no particular format restrictions exist. Also shown in Fig. 3 is a set of documents found in answer to the request on differential equations by using one of the available processing methods. The documents are listed in decreasing order of the correlation coefficient with the search request; a short 12-character identifier is shown for each document under the heading "answer," and full bibliographic citations are shown under "identification."

The average number of index terms used to identify each document is sometimes believed to be an important factor affecting retrieval performance. In the SMART system, this parameter is a difficult one to present and interpret, since the many procedures which exist for analyzing the documents and search requests generate indexing products with widely differing characteristics. A typical example is shown in Fig. 4, consisting of the index "vectors" generated by three different processing methods for the request on differential equations (short form "DIFFERNTL EQ"), and for document number 1 of the collection (short form "1A COMPUTER").

It may be seen from Fig. 4 that the number of terms identifying a document can change drastically from one method to another: for example, document number 1 is identified by 35 different word stems using the word stem analysis (labelled "null thesaurus" in Fig. 4); these 35 stems, however, give rise to 50 different concept numbers using the regular thesaurus, and to 55 concepts for the statistical phrase method. The number of index terms per document shown in the summary of Fig. 1 (35) must therefore be taken as an indication at best, and does not properly reflect the true situation.

In Fig. 4, each concept number is followed by some mnemonic characters to identify the concept and by a

* Practical considerations dictated the use of abstracts rather than full documents; the SMART system as such is not restricted to the manipulation of abstracts only.

```
*TEXT 2MICRO-PROGRAMMING .

$MICRO-PROGRAMMING
$R. J. MERCER (UNIVERSITY OF CALIFORNIA)
$U.S. GOV. RES. REPTS. VOL 30 PP 71-72(A) (AUGUST 15, 1958) PB 126893

MICRO-PROGRAMMING . THE MICRO-PROGRAMMING TECHNIQUE OF DESIGNING THE
CONTROL CIRCUITS OF AN ELECTRONIC DIGITAL COMPUTER TO FORMALLY INTERPRET
AND EXECUTE A GIVEN SET OF MACHINE OPERATIONS AS AN EQUIVALENT SET
OF SEQUENCES OF ELEMENTARY OPERATIONS THAT CAN BE EXECUTED IN ONE
PULSE TIME IS DESCRIBED .


*TEXT 3THE ROLE OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATIONS

$THE ROLE OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATIONS
$M. M. ASTRAHAN (IBM CORP.)
$IBM J. RES. AND DEV. VOL 2 PP 310-313 (OCTOBER 1958)

THE ROLE OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATIONS . THE ROLE
OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATIONS IS DISCUSSED . LARGE
MEMORIES PROVIDE AUTOMATIC REFERENCE TO MILLIONS OF WORDS OF MACHINE-RE-
ADABLE CODED INFORMATION OR TO MILLIONS OF IMAGES OF DOCUMENT PAGES
. HIGHER DENSITIES OF STORAGE WILL MAKE POSSIBLE LOW-COST MEMORIES
OF BILLIONS OF WORDS WITH ACCESS TO ANY PART IN A FEW SECONDS OR COMPLE-
TE SEARCHES IN MINUTES . THESE MEMORIES WILL SERVE AS INDEXES TO THE
DELUGE OF TECHNICAL LITERATURE WHEN THE PROBLEMS OF INPUT AND OF THE
AUTOMATIC GENERATION OF CLASSIFICATION INFORMATION ARE SOLVED . DOCUMENT
FILES WILL MAKE THE INDEXED LITERATURE RAPIDLY AVAILABLE TO THE SEARCHER
. MACHINE TRANSLATION OF LANGUAGE AND RECOGNITION OF SPOKEN INFORMATION
ARE TWO OTHER AREAS WHICH WILL REQUIRE FAST, LARGE MEMORIES .
```

FIG. 2. Typical Document Prints.

```
ANSWERS TO REQUESTS FOR DOCUMENTS ON SPECIFIED TOPICS          SEPTEMBER 28, 1964   PAGE 83

              CURRENT REQUEST - *LIST DIFFERNTL EQ NUMERICAL DIGITAL SOLN OF DIFFERENTIAL EQUATIONS

REQUEST   *LIST DIFFERNTL EQ NUMERICAL DIGITAL SOLN OF DIFFERENTIAL EQUATIONS
-------
          GIVE ALGORITHMS USEFUL FOR THE NUMERICAL SOLUTION OF ORDINARY
          DIFFERENTIAL EQUATIONS AND PARTIAL DIFFERENTIAL EQUATIONS ON DIGITAL
          COMPUTERS . EVALUATE THE VARIOUS INTEGRATION PROCEDURES (E.G. RUNGE--
          KUTTA, MILNE-S METHOD) WITH RESPECT TO ACCURACY, STABILITY, AND SPEED


ANSWER         CORRELATION         IDENTIFICATION
------         -----------         --------------
384STABILITY    0.6675       STABILITY OF NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS
                             W. E. MILNE AND R. R. REYNOLDS (OREGON STATE COLLEGE)
                             J. ASSOC. FOR COMPUTING MACH. VOL 6 PP 196-203 (APRIL, 1959)

ANSWER         CORRELATION         IDENTIFICATION
------         -----------         --------------
360SIMULATIN    0.5758       SIMULATING SECOND-ORDER EQUATIONS
                             D. G. CHADWICK (UTAH STATE UNIV.)
                             ELECTRONICS VOL 32 P 64 (MARCH 6, 1959)

ANSWER         CORRELATION         IDENTIFICATION
------         -----------         --------------
200SOLUTION     0.5663       SOLUTION OF ALGEBRAIC AND TRANSCENDENTAL EQUATIONS ON AN AUTOMATIC
                             DIGITAL COMPUTER
                             G.N. LANCE (UNIV. OF SOUTHAMPTON)
                             J. ASSOC. FOR COMPUTING MACH., VOL 6, PP 97-101, JAN., 1959
ANSWER         CORRELATION         IDENTIFICATION
------         -----------         --------------
392ON COMPUT    0.5508       ON COMPUTING RADIATION INTEGRALS
                             R. C. HANSEN (HUGHES AIRCRAFT CO.), L. L. BAILIN (UNIV. OF SOUTHERN
                             CALIFORNIA, AND R. W. RUTISHAUSER (LITTON INDUSTRIES, INC.)
                             COMMUN. ASSOC) FOR COMPUTING MACH. VOL 2 PP 28-31 (FEBRUARY, 1959)
ANSWER         CORRELATION         IDENTIFICATION
------         -----------         --------------
386ELIMINATI    0.5483       ELIMINATION OF SPECIAL FUNCTIONS FROM DIFFERENTIAL EQUATIONS
                             J. E. POWERS (UNIV. OF OKLAHOMA)
                             COMMUN. ASSOC. FOR COMPTING MACH. VOL 2 PP 3-4 (MARCH, 1959)
```

FIG. 3. Typical Search Request and Corresponding Answers.

```
OCCURRENCES OF CONCEPTS.AND PHRASES IN DOCUMENTS              SEPTEMBER 28, 1964

DOCUMENT       CONCEPT,OCCURS                                      PAGE 17

DIFFERNTL EQ   4EXACT 12  8ALGOR 12  13CALC 18  71EVAL   6  92DIGI 12
               110AUT 12  143UTI 12  176SOL 12  179STD 12  181QUA 24
               269ELI  4  274DIF 36  356VEL 12  357YAW  4  384TEG 12
               428STB  4  505APP 24

1A COMPUTER    2INPUT  4  5LOCAT 12  10ALPH 12  15BASE   6  16BASC  6
               31BIT   3  32REQU  3  41MCHO  8  47CHNG   6  53DATA  6   REGULAR
               57DSCB 15  59AMNT 24  72EXEC  6  77LIST   4  83MAP   6   THESAURUS
               87ENBL 12  930RDR 10  106NQU  6  107DGN 30  108LOO 12
               110AUT 36  1120PE  6  119AUT  8  121MEM   4  130MEA  4
               143UTI 12  146JOB 18  147SYS 12  149POG 36  158REL 12
               162ROF  6  163EAS 12  168ORD  4  176SOL 12  178SYM 18
               182SAV  4  187DIR 12  2100UT  4  212SIZ 12  216DOM 12
               276GEM 18  327AST 12  332SEE 12  338MCH   8  340LET  3
               346JET  6  350IFO  6  419GEM  6  501ORD   4  508ACT  6

DIFFERNTL EQ   ACCUR  12  ALGORI 12  COMPUT 12  DIFFER 24  DIGIT  12
               EQU    24  EVALU  12  GIVE   12  INTEGR 12  METHOD 12
               NUMER  12  ORDIN  12  PARTI  12  PROCED 12  RUNGE- 12
               SOLUT  12  SPEED  12  STABIL 12  USE    12  VARIE  12
                                                                        NULL
1A COMPUTER    BAS    12  CHARAC 12  COMPUT 36  DESCRI 12  DESIGN 12   THESAURUS
               DIRECT 12  ENABLE 12  ESTIM  12  EXPLAI 12  FORM   12
               GIVE   12  HANDLE 12  ICLUST 12  INDEPE 12  INFORM 12
               MACHIN 24  OPER   12  ORD    12  ORIENT 12  PLANE  12
               POS    12  POSS   12  PROBLE 36  PROGRA 36  RECOGN 12
               SCANN  12  SIMPLE 12  SIZE   24  STORE  12  STRUCT 12
               TECHNI 12  TOWARD 12  TRANSF 12  USING  12  WRITT  12

DIFFERNTL EQ   4EXACT 12  8ALGOR 12  13CALC 18  71EVAL   6  92DIGI 12
               110AUT 12  143UTI 12  176SOL 12  179STD 12  181QUA 24
               269ELI  4  274DIF 36  356VEL 12  357YAW  4  375NUM 36
               379DIF 72  384TEG 12  428STB  4  505APP 24

1A COMPUTER    2INPUT  4  5LOCAT 12  10ALPH 12  14COOR 72  15BASE  6  STAT.       STAT.
               16BASC  6  31BIT   3  32REQU  3  41MCHO  8  47CHNG  6  PHRASES     PHRASE
               53DATA  6  57DSCB 15  59AMNT 24  72EXEC  6  77LIST  4              LOOK-UP
               83MAP   6  87ENBL 12  930RDR 10  106NQU  6  107DGN 30
               108LOO 12  110AUT 36  1120PE  6  119AUT  8  121MEM  4
               130MEA  4  143UTI 12  146JOB 18  147SYS 12  149POG 36
               158REL 12  162RUF  6  163EAS 12  168ORD  4  176SOL 12
               178SYM 18  182SAV  4  187DIR 12  200DA- 72  2100UT  4
               212SIZ 12  21600M 12  219POG 36  276GEM 18  292THK 36
               302LOO 72  327AST 12  332SEE 48  338MCH  8  340LET  3
               346JET  6  350IFO  6  419GEM  6  501ORD  4  508ACT  6
```

FIG. 4. Typical Indexing Products for Three Analysis Procedures.

weight. The weights assigned to the concept numbers also change from method to method. Since no distinction is made in the evaluation procedure between retrieved and nonretrieved documents, the last indicator included in Fig. 1 (the number of retrieved documents per request) must also be put into the proper perspective. A discussion of this point is postponed until after the evaluation measures are introduced in the next few paragraphs.

• Evaluation Measures

1. *Recall and Precision*

One of the most crucial tasks in the evaluation of retrieval systems is the choice of measures which reflect systems performance. In the present context, such a measurement must of necessity depend primarily on the system's ability to retrieve wanted information and to reject nonwanted material, to the exclusion of operational criteria such as retrieval cost, waiting time, input preparation time, and so on. The last mentioned factors may be of great practical importance in an operational situation, but do not enter, at least initially, into the evaluation of experimental procedures.

A large number of measures have been proposed in the past for the evaluation of retrieval performance.[4] Perhaps the best known of these are, respectively, *recall* and *precision; recall* is defined as the proportion of relevant material actually retrieved, and *precision* as the proportion of retrieved material actually relevant.* A system with high recall is one which rejects very little that is relevant but may also retrieve a large proportion of irrelevant material, thereby depressing precision. High precision, on the other hand, implies that very little irrelevant information is produced but much relevant information may be missed at the same time, thus depressing recall. Ideally, one would of course hope for both high recall and high precision.†

Measures such as recall and precision are particularly attractive when it comes to evaluating *automatic* retrieval procedures, because a large number of extraneous factors which cause uncertainty in the evaluation of conventional (manual) systems are automatically absent. The following characteristics of the present system are particularly important in this connection:

(a) input errors in the conventional sense, due to faulty indexing or encoding, are eliminated since all indexing operations are automatic;

* Precision has also been called "relevance," notably in the literature of the ASLIB–Cranfield project.[5]

† It has, however, been conjectured that an inverse relationship exists between recall and precision, such that high recall automatically implies low precision and vice versa.

(b) for the same reason, conventional search errors arising from the absence of needed search terms are also excluded;
(c) errors cannot be introduced in any transition between original search request and final machine query, since this transition is now handled automatically and becomes indistinguishable from the main analysis operation;
(d) inconsistencies introduced by a large number of different indexers and by the passage of time in the course of an experiment cannot arise; and
(e) the role of human memory as a disturbance in the generation of retrieval measurements is eliminated (this factor can be particularly troublesome when source documents are to be retrieved in a conventional system by persons who originally perform the indexing task).

In order to calculate the *standard* recall and precision measures the following important tasks must be undertaken:

(a) relevance judgments must be made by hand in order to decide, for each document and for each search request, whether the given document is relevant to the given request;
(b) the relevance judgments are usually all or nothing decisions so that a given document is assumed either wholly relevant or wholly irrelevant (in case of doubt relevance is assumed); and
(c) a cut-off in the correlation between documents and search requests is normally chosen, such that documents whose correlation exceeds the cut-off value are retrieved, while the others are not retrieved.

2. *The Generation of Relevance Judgments*

A great deal has been written concerning the difficulties and the appropriateness of the various operations listed in part 1.[5–8] The first task, in particular, which may require the performance of hundreds of thousands of human relevance judgments for document collections of reasonable size, is extremely difficult to satisfy and to control.

Two solutions have been suggested, each of which would base the relevance decisions on less than the whole document collection. The first one consists in using sampling techniques to isolate a suitable document subset, and in making relevance judgments only for documents included in that subset. If the results obtained for the subset, however, are to be applicable to the total collection, it becomes necessary to choose a sample representative of the whole. For most document collections, this turns out to be a difficult task.

The other solution consists in formulating search requests based on specific source documents included in the collection, and in measuring retrieval performance for a given search request as a function of the retrieval of the respective source documents. This procedure suffers from the fact that search requests based on source documents are often claimed to be nontypical, thus introducing a bias into the measurements which does not exist for requests reflecting actual user needs.

Since the document collection used in connection with the present experiments is small enough to permit an exhaustive determination of relevance, the possible pitfalls inherent in the sampling procedure and in the use of source documents were avoided to a great extent. Many of the problems connected with the rendering of relevance judgments are, however, unresolved for general document collections.

3. *The Cut-off Problem*

The other major problem is caused by the requirement to pick a correlation cut-off value to distinguish retrieved documents from those not retrieved. Such a cut-off introduces a new variable which seems to be extraneous to the principal task of measuring retrieval performance. Furthermore, in the SMART system, a different cut-off would have to be picked for each of the many processing methods if it were desired to retrieve approximately the same number of documents in each case.

Because of these added complications, it was felt that the standard recall and precision measures should be redefined so as to remove the necessary distinction between retrieved and nonretrieved information. Fortunately, this is not difficult in computer-based information systems, because in such systems numeric coefficients expressing the similarity between each document and each search request are obtained as output of the search process. Documents may then be arranged in decreasing order of these similarity coefficients, as shown, for example, for the previously used request on differential equations in the center section of Fig. 5. It may be seen in the figure that document 384 exhibits the largest correlation with the search request, followed by documents 360, 200, 392, and so on.

An ordered document list of the kind shown in Fig. 5 suggests that a suitable criterion for recall and precision measures would be the set of *rank-orders* of the *relevant* documents, when these documents are arranged in decreasing correlation order. A function of the rank-order list which penalizes high ranks for relevant documents (and therefore low correlation coefficients) can be used to express recall, while a function penalizing low ranks of nonrelevant documents is indicative of precision.

4. *Normalized Recall and Normalized Precision**

It is desired to use as a measure of retrieval effectiveness a set of parameters which reflects the standard recall and the standard precision, and does not depend on a distinction between retrieved and nonretrieved documents. This suggests that one might take the average of the recall and the average of the precision obtained for

* The measures described in this part were suggested by J. Rocchio.[9]

ROW CORRELATIONS OF PHRASE-DOCUMENT MATRIX   SEPTEMBER 28, 1964   PAGE 73

a) INCREASING DOCUMENT ORDER			
DIFFERNTL	EQ	1A COMPUTER	0.1234
DIFFERNTL	EQ	2MICRO-PROGR	0.0875
DIFFERNTL	EQ	3THE ROLF OF	0.0293
DIFFERNTL	EQ	4A NEW CLASS	0.0844
DIFFERNTL	EQ	5ANALYSIS OF	0.0658
DIFFERNTL	EQ	6GENERALIZED	0.0741
DIFFERNTL	EQ	7AN IMPROVED	0.2090
DIFFERNTL	EQ	8SHORT-CUT M	0.0861
DIFFERNTL	EQ	9OPERATION A	0.0611
DIFFERNTL	EQ	10ACCURATE T	0.1103
DIFFERNTL	EQ	12DIGITAL CO	0.0883
DIFFERNTL	EQ	13HALF-ADDER	0.0548
DIFFERNTL	EQ	16CONTROL AP	0.0336
DIFFERNTL	EQ	17THE FUNCTI	0.0583
DIFFERNTL	EQ	18AN ACCURAT	0.1397
DIFFERNTL	EQ	19RESISTANCE	0.0177
DIFFERNTL	EQ	20DIFFERENTI	0.2123
DIFFERNTL	EQ	21AN ERROR-C	0.2105
DIFFERNTL	EQ	22LATCHING C	0.0057
DIFFERNTL	EQ	23MINIATURE	0.0307
DIFFERNTL	EQ	24SOME NOVEL	0.0199
DIFFERNTL	EQ	25A NEW TRAN	0.1068
DIFFERNTL	EQ	26SEMICONDUC	0.0653
DIFFERNTL	EQ	27TEN MEGAPU	0.1004
DIFFERNTL	EQ	28DESIGN OF	0.1375
DIFFERNTL	EQ	29INVESTIGAT	0.0879
DIFFERNTL	EQ	30A TRANSIST	0.0736
DIFFERNTL	EQ	31MAGNETIC C	0.0575
DIFFERNTL	EQ	32ANALOGUE I	0.2283
DIFFERNTL	EQ	33THE USE OF	0.0802
DIFFERNTL	EQ	34END-FIRED	0.0458
DIFFERNTL	EQ	35A LOAD-SHA	0.0331
DIFFERNTL	EQ	36FUNDAMENTA	0.0392
DIFFERNTL	EQ	37A HIGH-SPE	0.0364
DIFFERNTL	EQ	38AUTOMATIC	0.1043
DIFFERNTL	EQ	41COMMUNICAT	0.1185
DIFFERNTL	EQ	42A DIRECT R	0.0439
DIFFERNTL	EQ	43THE DATA C	0.0333
DIFFERNTL	EQ	44ACCURACY C	0.1399
DIFFERNTL	EQ	45A CALCULAT	0.2958
DIFFERNTL	EQ	46RADIO DIRE	0.0980
DIFFERNTL	EQ	47SPECIAL PU	0.1268
DIFFERNTL	EQ	48A BUSINESS	0.0086
DIFFERNTL	EQ	49A DUAL MAS	0.0575
DIFFERNTL	EQ	50ACCURACY C	0.0668
DIFFERNTL	EQ	52+ATHENA ,	0.1030
DIFFERNTL	EQ	53A COMPUTER	0.1327
DIFFERNTL	EQ	54AN AUTOMAT	0.0763
DIFFERNTL	EQ	55AUTOMATIC	0.0746
DIFFERNTL	EQ	56THE COMPUT	0.1513
DIFFERNTL	EQ	57CASE STUDY	0.0950
DIFFERNTL	EQ	58THE LARGES	0.0256
DIFFERNTL	EQ	59DATA PROCE	0.0302
DIFFERNTL	EQ	60INTELLIGEN	0.0091
DIFFERNTL	EQ	61AN INPUT R	0.0404
DIFFERNTL	EQ	62ON PROGRAM	0.1181

b) DECREASING CORRELATION ORDER			
DIFFERNTL	EQ	384STABILITY	0.6675
DIFFERNTL	EQ	360SIMULATIN	0.5758
DIFFERNTL	EQ	200SOLUTION	0.5663
DIFFERNTL	EQ	392ON COMPUT	0.5508
DIFFERNTL	EQ	386ELIMINATI	0.5483
DIFFERNTL	EQ	103RUNGE-KUT	0.5444
DIFFERNTL	EQ	85NOTE ON AN	0.4510
DIFFERNTL	EQ	192SOLVING E	0.4106
DIFFERNTL	EQ	358STABILIZA	0.3986
DIFFERNTL	EQ	102ON THE SO	0.3986
DIFFERNTL	EQ	387BOUNDARY	0.3968
DIFFERNTL	EQ	202STABLE PR	0.3906
DIFFERNTL	EQ	229MATRIX PR	0.3505
DIFFERNTL	EQ	88PROPOSED M	0.3451
DIFFERNTL	EQ	251ERROR EST	0.3329
DIFFERNTL	EQ	234ANALOGUE	0.3176
DIFFERNTL	EQ	253ROUND-OFF	0.3152
DIFFERNTL	EQ	186ALGORITHM	0.3144
DIFFERNTL	EQ	169THEORETIC	0.3136
DIFFERNTL	EQ	128COMPUTER	0.3034
DIFFERNTL	EQ	226+DEPI ..	0.3028
DIFFERNTL	EQ	45A CALCULAT	0.2958
DIFFERNTL	EQ	390MONTE CAR	0.2866
DIFFERNTL	EQ	388A METHOD	0.2787
DIFFERNTL	EQ	173AUTOMATIC	0.2753
DIFFERNTL	EQ	306ELECTRONI	0.2750
DIFFERNTL	EQ	318FROM FORM	0.2741
DIFFERNTL	EQ	249MATHEMATI	0.2683
DIFFERNTL	EQ	266UNIFYING	0.2682
DIFFERNTL	EQ	217SIMULATIO	0.2665
DIFFERNTL	EQ	367ON EXPONE	0.2661
DIFFERNTL	EQ	213PREDICTIO	0.2630
DIFFERNTL	EQ	108SECANT MO	0.2620
DIFFERNTL	EQ	383A NOTE ON	0.2580
DIFFERNTL	EQ	191DIGITAL C	0.2370
DIFFERNTL	EQ	171SMALL COM	0.2325
DIFFERNTL	EQ	248METHOD FO	0.2319
DIFFERNTL	EQ	283BINARY AR	0.2311
DIFFERNTL	EQ	252A CLASS O	0.2303
DIFFERNTL	EQ	385NUMERICAL	0.2300
DIFFERNTL	EQ	210EVALUATIO	0.2285
DIFFERNTL	EQ	220DATA PREP	0.2282
DIFFERNTL	EQ	32ANALOGUE I	0.2282
DIFFERNTL	EQ	197TECHNICAL	0.2280
DIFFERNTL	EQ	355A ROUTINE	0.2272
DIFFERNTL	EQ	215DIGITAL C	0.2259
DIFFERNTL	EQ	69COMPUTERS	0.2249
DIFFERNTL	EQ	201ITERATIVE	0.2198
DIFFERNTL	EQ	193ARTIFICIA	0.2196
DIFFERNTL	EQ	361SAINT COM	0.2187
DIFFERNTL	EQ	257SURVEY OF	0.2181
DIFFERNTL	EQ	236OPERATING	0.2180
DIFFERNTL	EQ	117COMPUTATI	0.2170
DIFFERNTL	EQ	207AN APPLIC	0.2162
DIFFERNTL	EQ	200IFFERENTI	0.2122
DIFFERNTL	EQ	235FREEZING	0.2093

c) HISTOGRAM	
0.9800	0
0.9600	0
0.9400	0
0.9200	0
0.9000	0
0.8800	0
0.8600	0
0.8400	0
0.8200	0
0.8000	0
0.7800	0
0.7600	0
0.7400	0
0.7200	0
0.7000	0
0.6800	0
0.6600	1
0.6400	1
0.6200	1
0.6000	1
0.5800	1
0.5600	3
0.5400	6
0.5200	6
0.5000	6
0.4800	6
0.4600	6
0.4400	7
0.4200	7
0.4000	8
0.3800	12
0.3600	12
0.3400	14
0.3200	15
0.3000	21
0.2800	23
0.2600	33
0.2400	34
0.2200	47
0.2000	60
0.1800	73
0.1600	87
0.1400	106
0.1200	135
0.1000	166
0.0800	200
0.0600	257
0.0400	304
0.0200	348

FIG. 5. Correlations Between Search Request and Document Collection.

all possible retrieval levels to define a new pair of measures, termed respectively *normalized* recall and *normalized* precision. Specifically, if $R_{(j)}$ is the standard recall after retrieving j documents from the collection (that is, if $R_{(j)}$ is equal to the number of relevant documents retrieved divided by the total relevant in the collection, assuming j documents retrieved in all), then the normalized recall can be defined as

$$R_{\text{norm}} = \frac{1}{N}\sum_{j=1}^{N} R_{(j)}$$

where N is the total number of documents in the collection.

Similarly, if $P_{(j)}$ is the standard precision after retrieving j documents from the collection, then a normalized precision measure is defined as

$$P_{\text{norm}} = \frac{1}{N}\sum_{j=1}^{N} P_{(j)} .$$

R_{norm} and P_{norm} may thus be obtained mechanically by first retrieving one document and calculating recall and precision, then retrieving another document, and again calculating R and P, and repeating the process one document at a time until all documents in the whole collection have been retrieved. Finally, all R's and P's are averaged to obtain the normalized measures.

In practice, this manner of proceeding would be extremely tedious for large document collections, even if the calculations were done by computer. It may, however, be shown by reasonably straightforward algebra that the normalized recall and normalized precision may be rewritten, respectively, as

$$R_{\text{norm}} = \frac{1}{N}\sum_{j=1}^{N} R_{(j)} = 1 - \frac{\sum_{i=1}^{n} r_i - \sum_{i=1}^{n} i}{n(N-n)} \qquad (1)$$

and

$$P_{\text{norm}} = \frac{1}{N}\sum_{j=1}^{N} P_{(j)} = 1 - \frac{\sum_{i=1}^{n} ln\, r_i - \sum_{i=1}^{n} ln\, i}{ln \dfrac{N!}{n!(N-n)!}} \qquad (2)$$

where r_i is the rank (in decreasing correlation order with the search request) of the ith relevant document in the collection,

n is the total number of relevant documents in the collection,

and N is the total number of documents in all.

The expressions in the right-hand side are suitable for automatic computation and are in fact used in the SMART system. All basic definitions are summarized in Fig. 6.

	STANDARD DEFINITIONS (BASED ON THRESHOLD TO DISTINGUISH DOCUMENTS RETRIEVED FROM DOCUMENTS NOT RETRIEVED)	DEFINITIONS BASED ON RANKS OF RELEVANT DOCUMENTS (DOCUMENTS ARRANGED IN ORDER BY DECREASING CORRELATION WITH SEARCH REQUESTS)
RECALL	PROPORTION OF RELEVANT MATERIAL ACTUALLY RETRIEVED (LOW RECALL IMPLIES THAT SOME RELEVANT DOCUMENTS HAVE LOW CORRELATION WITH SEARCH REQUEST AND ARE THUS NOT RETRIEVED)	$1 - \frac{\sum_{i=1}^{n} r_i - \sum_{i=1}^{n} i}{n(N-n)}$ n: NUMBER OF RELEVANT DOCUMENTS N: NUMBER OF DOCUMENTS IN COLLECTION r_i: RANK ORDER OF i^{th} RELEVANT DOCUMENT
PRECISION	PROPORTION OF RETRIEVED MATERIAL ACTUALLY RELEVANT (LOW PRECISION IMPLIES THAT SOME NON-RELEVANT DOCUMENTS HAVE A HIGH CORRELATION WITH SEARCH REQUEST AND ARE THUS RETRIEVED)	$1 - \frac{\sum_{i=1}^{n} \ln r_i - \sum_{i=1}^{n} \ln i}{\ln \frac{N!}{n!(N-n)!}}$

Fig. 6. Basic Definitions of Recall and Precision.

```
                                      SEPTEMBER 28, 1964   PAGE 86
EVALUATION OF REQUEST DIFFERNTL EQ WITH  16 RELEVANT DOCUMENTS

   THE TOP FIFTEEN DOCUMENTS           RELEVANT DOCUMENT RANKS

    1 X 384STABILITY 0.6676             1 384STABILITY 0.6676
    2 X 360SIMULATIN 0.5758             2 360SIMULATIN 0.5758
    3 X 200SOLUTION  C.5664             3 200SOLUTION  0.5664
    4 X 392ON COMPUT 0.5508             4 392ON COMPUT 0.5508
    5 X 386ELIMINATI 0.5484             5 386ELIMINATI 0.5484
    6 X 103RUNGE-KUT 0.5445             6 103RUNGE-KUT 0.5445    REGULAR
    7 X 85NOTE ON AN C.4511             7 85NOTE ON AN 0.4511    THESAURUS
    8   192SOLVING E 0.4106             9 102ON THE SO 0.3987
    9 X 102ON THE SO 0.3987            10 358STABILIZA 0.3986
   10 X 358STABILIZA C.3986            11 387BOUNDARY  0.3968
   11 X 387BOUNDARY  0.3968            12 202STABLE PR 0.3907
   12 X 202STABLE PR 0.3907            15 251ERROR EST 0.3329
   13   229MATRIX PR 0.3506            17 253ROUND-OFF 0.3152
   14   88PROPOSED M 0.3452            23 390MONTE CAR 0.2866
   15 X 251ERROR EST C.3329            24 388A METHOD  0.2788
                                       40 385NUMERICAL 0.2301

RANK RECALL= 0.7196 LOG PRECISION= 0.9169
NORMALIZED RECALL=0.9914626 NORMALIZED PRECISION=0.9572
RNK REC + LOG PRE=1.6365 WEIGHTED NORMED RECALL + NORMED PREC=1.9146

   THE TOP FIFTEEN DOCUMENTS           RELEVANT DOCUMENT RANKS

    1 X 384STABILITY 0.8576             1 384STABILITY 0.8576
    2 X 360SIMULATIN 0.7741             2 360SIMULATIN 0.7741
    3 X 386ELIMINATI 0.7408             3 386ELIMINATI 0.7408
    4 X 392ON COMPUT 0.6571             4 392ON COMPUT 0.6571
    5 X 200SOLUTION  0.6444             5 200SOLUTION  0.6444
    6 X 85NOTE ON AN 0.6372             6 85NOTE ON AN 0.6372    STATISTICAL
    7 X 387BOUNDARY  0.6072             7 387BOUNDARY  0.6072    PHRASE SEARCH
    8 X 103RUNGE-KUT C.5875             8 103RUNGE-KUT 0.5875
    9 X 102ON THE SO 0.5648             9 102ON THE SO 0.5648
   10 X 390MONTE CAR 0.5448            10 390MONTE CAR 0.5448
   11 X 358STABILIZA 0.5437            11 358STABILIZA 0.5437
   12 X 388A METHOD  C.5318            12 388A METHOD  0.5318
   13 X 202STABLE PR 0.5163            13 202STABLE PR 0.5163
   14 X 385NUMERICAL 0.4942            14 385NUMERICAL 0.4942
   15   169THEORETIC 0.4794            21 251ERROR EST 0.3444
                                       25 253ROUND-OFF 0.3157

RANK RECALL= 0.9007 LOG PRECISION= 0.9751
NORMALIZED RECALL=0.9975900 NORMALIZED PRECISION=0.9880
RNK REC + LOG PRE=1.8758 WEIGHTED NORMED RECALL + NORMED PREC=1.9759
```

Fig. 7. Automatic Evaluation.

• Test Results

1. *Output Formats*

The normalized recall and precision measures are a function only of the ranks of the relevant documents. If those measures are to be evaluated automatically as part of the retrieval process, it is necessary to introduce for each search request processed a list of the corresponding relevant document identifications. To this effect the requester is given a copy of the full document collection *after* his request is received, and he is asked to list those documents which he believes should be considered relevant to his request. It is important to note that these relevance judgments are *a priori* judgments, based on the document texts only, and not on any retrieval results produced by the computer.

The type of output obtained from the evaluation process is illustrated in Fig. 7. The top part of the figure represents the output from the regular thesaurus procedure for the request on differential equations previously used, while the bottom part is produced by the statistical phrase method. On the right side of the figure appears the list of all 16 relevant document numbers, as originally submitted by the user, together with the respective correlation coefficients and the ranks assigned by the computer during the retrieval process. It may be noticed that the relevant document which exhibits the lowest correlation with the search request is ranked 40th out of 405 by the regular thesaurus procedure, but only 25th out of 405 by the statistical phrase search.

The document ranks are used by the program to produce a variety of measures reflecting recall and precision, including the normalized recall and normalized precision measures previously introduced. Also calculated are simplified expressions, termed respectively *rank recall* and *log precision,* and defined as follows:

$$\text{rank recall} = \frac{\sum_{i=1}^{n} i}{\sum_{i=1}^{n} r_i}$$

$$\text{log precision} = \frac{\sum_{i=1}^{n} \ln i}{\sum_{i=1}^{n} \ln r_i}\,.$$

These simple measures are analogous to the normalized recall and normalized precision but do not take into account the collection size N.

Finally, two composite measures are produced which include both recall and precision components. The first one consists simply of the sum of rank recall plus log precision. The other is a weighted sum of the normalized measures, as follows:

$$\text{normed overall measure} = 1 - 5\,(R_{\text{norm}}) + P_{\text{norm}}$$

The factor of 5 is so chosen as to give equal weight to the two component measures.

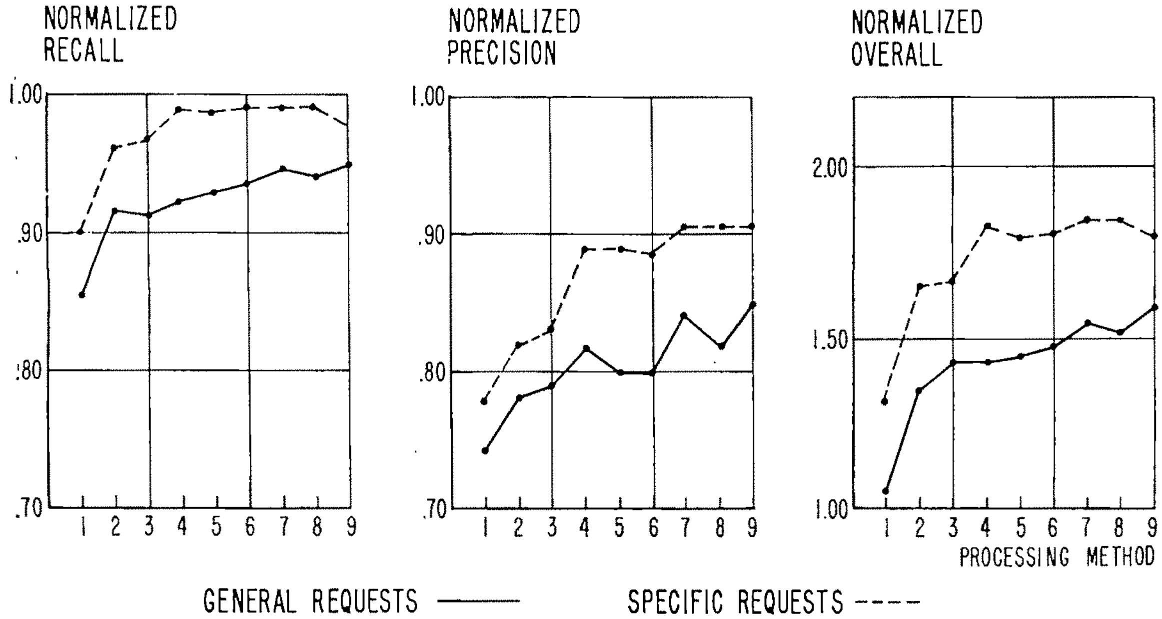
Also included on the left-hand side of Fig. 7 are lists of the 15 documents which exhibit the highest correlation coefficients with the search request. The relevant documents on that list are provided with a special marker (X). It may be seen for the example of Fig. 7 that the recall and precision values obtained by the statistical phrase process are larger than the corresponding values for the thesaurus lookup procedure.

2. Results Derived from the Normalized Measures

In order to obtain statistically useful measurements, the recall and precision values must be averaged over many different search requests. This is done in Fig. 8 for nine different processing methods and for a total of ten specific and seven general requests.

The following processing methods are included in Fig. 8:

1. Thesaurus—Titles only
   The word stems included in the titles of the documents are looked up in the regular thesaurus and replaced by weighted concept numbers. The remainder of the document abstracts is not used.
2. Thesaurus—Hierarchy (up and add)
   The complete document abstracts are used. All word stems are replaced by weighted concept numbers from the thesaurus; these concept numbers are then looked up in the hierarchy, and to each original concept the corresponding "parent" from the next higher hierarchy level is added.
3. Thesaurus—Logical Vectors
   Complete document abstracts are used. All word stems are replaced by concept numbers from the thesaurus, and each concept is given a weight of 1.
4. Word Stems—Full Text (Null Thesaurus)
   Complete document abstracts are used, and weighted word stems are generated by the suffix cut-off procedure. No further dictionary is used.
5. Thesaurus—Syntactic Phrases
   The weighted concepts obtained from the thesaurus are looked up in the phrase dictionary, and phrase concepts corresponding to available concept groupings are detected and used as document identifiers, provided that certain specified syntactic relationships hold between the phrase components.
6. Thesaurus—Hierarchy (down and add)
   Procedure identical with 2 except that the concepts added from the hierarchy are obtained by



PROCESSING METHODS:

1. THESAURUS – TITLES ONLY
2. THESAURUS – HIERARCHY (UP AND ADD)
3. THESAURUS – LOGICAL VECTORS
4. WORD STEMS – FULL TEXT
5. THESAURUS – SYNTACTIC PHRASES
6. THESAURUS – HIERARCHY (DOWN AND ADD)
7. THESAURUS – NUMERIC VECTORS
8. THESAURUS – STATISTICAL PHRASES (REQUESTS ONLY)
9. THESAURUS – STATISTICAL PHRASES (WHOLE)

Fig. 8. Normalized Recall, Precision and Overall Measures (averaged over 10 specific and 7 general requests for several processing methods).

taking the "sons" of each original concept on the next lower level of the hierarchy.

7. Thesaurus—Numeric Vectors
Procedure identical with 3 except that the concepts obtained from the thesaurus are weighted in accordance with their frequency. This is the standard thesaurus method.

8. Thesaurus—Statistical Phrases (requests only)
Standard thesaurus method (7), to which are added phrases (concept groupings) occurring in the requests only. Syntactic relationships between phrase components are not used.

9. Thesaurus—Statistical Phrases (whole)
Standard thesaurus method (7) followed by the phrase procedure (5) for all documents without detection of syntactic relationships between phrase components.

The data of Fig. 8 give rise to the following observations:

(a) the normalized evaluation measures obtained for the various processing methods exhibit substantial differences;

(b) as one proceeds from one method to another, both recall and precision tend to vary in the same direction (either up or down);

(c) all the measures (recall, precision, and overall) obtained for the specific requests are larger than the corresponding values for the general requests, thus indicating a better systems performance for clearly specified topic classes*;

(d) methods one to four tend to produce relatively poorer recall than methods five to nine; these same methods also furnish relatively poor precision;

(e) the use of the regular thesaurus which provides vocabulary control (method seven) seems much more effective than the use of the original words included in document and search requests (method four)†;

(f) the most effective procedures seem to be those which use combinations of concepts (phrases), rather than individual concepts alone.

The data of Fig. 8 are of interest in themselves, since they do support the notion that more reasonable procedures (than mere word matching) can be generated to improve retrieval effectiveness in an automatic system. However, if full advantage is to be taken of the organization of the SMART system, then search requests are best processed iteratively by several different methods, and the respective outputs combined. In order to determine whether this juxtaposition of methods can in fact be used to improve the performance characteristics, average normalized recall and precision figures are given in Fig. 9 for six *combined methods* and for the 17 requests previously used in Fig. 8.

Figure 9 includes the normalized recall and precision values for the regular thesaurus run previously shown in Fig. 8, followed by the same measures for various combined methods. All of the combined runs include the regular thesaurus run (method 7 of Fig. 8) as a component. In fact, the following correspondences between Figs. 8 and 9 are apparent:

Method 1 of Fig. 9	corresponds to method 7	of Fig. 8,
method 2 " "	corresponds to methods 7 + 4	" "
method 3 " "	corresponds to methods 7 + 9	" "
method 4 " "	corresponds to methods 7 + 6	" "
method 5 " "	corresponds to methods 7 + 2	" "
method 6 " "	corresponds to methods 7 + 4 + 9	" "

It may be seen that for three of the combined methods of Fig. 8 (methods 2, 3, and 6), the overall measures for both specific and general requests are larger than for any of the included methods alone. Method 6, consisting of a combination of regular thesaurus plus word stems plus statistical phrase runs, seems to be particularly effective.

The normalized recall and precision measures for the combined methods are computed by using the rank lists produced by the computer for the individual methods alone, and automatically generating a *combined rank list*. The combined rank of a given document depends on the individual ranks held by that document in the component methods. Specifically, documents are taken alternately from the component lists to form the new combined list, and a document already included on the combined list is rejected if an attempt is made to list it again. The final combined rank list is then used to compute normalized recall and precision measures for the combined methods, as specified in the previous section. The resulting measures are averaged over several search requests to produce the graphs of Fig. 9.

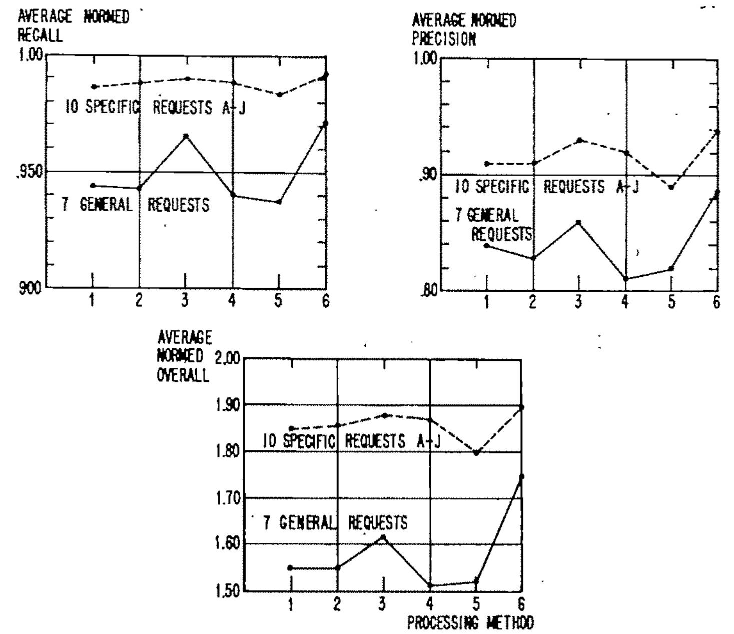
A combined rank list, generated for the two methods illustrated by the evaluation output of Fig. 7, is shown in Fig. 10 (only the first 15 documents are included for each component method). Documents previously specified as relevant are marked with an X, as in Fig. 7.

3. *Results Using the Standard Measures*

The performance characteristics of the SMART retrieval operations are reflected with reasonable accuracy in the data of Figs. 8 and 9. In particular, these figures

* These results would seem to indicate that Cleverdon's observations reported by Swet[4], that specific requests will have high precision and low recall and vice versa for general requests, need not necessarily hold in all circumstances.

† This observation has of course been made many times before, particularly by librarians and documentalists, but still requires emphasis in computer circles.



PROCESSING METHODS:

1. REGULAR THESAURUS
2. REGULAR THESAURUS + WORD STEMS
3. REGULAR THESAURUS + STATISTICAL PHRASES
4. REGULAR THESAURUS + HIERARCHY (DOWN)
5. REGULAR THESAURUS + HIERARCHY (UP)
6. REGULAR THESAURUS + WORD STEMS + STATISTICAL PHRASES

Fig. 9. Normalized Recall, Precision and Overall Measures for Several Merged Methods.

can be used to obtain an idea of the *relative* effectiveness of one method compared with another. The data are, however, difficult to interpret in absolute terms, particularly since the measures used are new ones, and no comparable output is available elsewhere in the literature.

In order to furnish some indication of systems performance which could lend itself more easily to a comparison with previously published data, the *standard* recall and precision measures, reflecting respectively the proportion of relevant material retrieved and the proportion of retrieved material relevant, are also computed for the search requests previously used. To generate these functions, it becomes necessary to choose appropriate threshold values which separate the retrieved information from that not retrieved. The procedure adopted for this purpose is as follows:

(a) a specified standard recall value is picked (say, 0.1);

(b) the number of documents which must be retrieved for a given search request in order to produce the specified recall is determined;

(c) using the cut-off value calculated under (b) for the number of retrieved documents, the precision measure (corresponding to the specified recall) is generated;

(d) the precision values obtained for a given recall level are averaged over a number of search requests, and the corresponding point is plotted on a precision versus recall plot;

(e) the complete procedure is repeated for a new recall level (say, 0.2, and 0.3, and so on) to produce a curve of the type shown in Fig. 11.

Figure 11 displays the standard precision versus standard recall graphs obtained for six processing methods, averaged over the 17 search requests previously used in Figs. 8 and 9. Figure 11 is in the exact form introduced by Cleverdon,[5,6] using the standard precision and recall measures, rather than the normalized measures based on rank lists. The procedure described above to generate the average precision over several requests is, however, different from Cleverdon's, since he combines requests not by computing separate recall-precision points for each request which are then averaged, but rather by treating sets of requests with $i,j,...,k$ relevant documents, respectively, as a single request with $i + j + ... + k$ relevant documents in all. Although the actual measurements are thus conducted from a somewhat different point of view, the output plots presented here should, nevertheless, lend themselves to a comparison with the published Cranfield material.*

* Recall versus precision plots have been criticized, because important information reflected in separate plots of recall and precision is obscured in the combined presentation (notably the number of documents both retrieved and relevant.)[10]

Ranks		Document Numbers	Ranks		Document Numbers	Ranks		Document Numbers
1	X	384	1	X	384	1	X	~~384~~
2	X	360	2	X	360	2	X	~~360~~
3	X	200	3	X	200	3	X	386
4	X	392	4	X	386	4	X	~~392~~
5	X	~~386~~	5	X	392	5	X	~~200~~
6	X	103	6	X	103	6	X	85
7	X	~~85~~	7	X	85	7	X	387
8		192	8	X	387	8	X	~~103~~
9	X	102	9		192	9	X	~~102~~
10	X	358	10	X	102	10	X	390
11	X	~~387~~	11	X	358	11	X	~~358~~
12	X	202	12	X	390	12	X	388
13		229	13	X	202	13	X	~~202~~
14		88	14	X	388	14	X	385
15	X	251	15		229	15		169
	⋮		16		88		⋮	
			17	X	385			
			18	X	251			
			19		169			
				⋮				

(a) Regular Thesaurus (b) Combined Rank List (c) Statistical Phrases

FIG. 10. Merging of Rank-order Lists.

The data of Fig. 11 confirm those previously shown in Fig. 8 in that the statistical phrase run again seems to give the best performance. Furthermore, word stem comparisons are again inferior to the regular thesaurus runs, and "titles only" analysis is generally inferior. The differences in systems performance previously noted for the output of Figs. 8 and 9 are again in evidence, since, for a given recall level, average precision can vary by over 30 percent from one method to another. The same is true of the average recall differences for a given level of precision.

Figure 12 shows standard precision versus standard recall figures averaged separately over the specific and the general requests for three processing methods. A comparison with Fig. 9 again indicates that both recall and precision measures are substantially higher for the specific requests than for the general requests.

• Conclusions

The evaluation procedures and results included in the present study are based on the manipulation of one relatively small collection of document abstracts, and a set of about 20 search requests. Only about 15 different processing methods are used. Under the circumstances, it is not possible to make claims of general validity, or to prove many assertions with finality.

Nevertheless, it is believed that the data presented here can be used as indications of the kind of performance to be expected of automatic retrieval systems. In particular, the data which point to the existence of considerable discrepancies in performance characteristics between processing methods may be expected to be confirmed by new experiments with different document collections and larger numbers of search requests. Of special interest, in this connection, is the fact that certain processing methods exhibit *both* high recall and high precision, thus indicating good overall performance.

The other principal piece of evidence tends to support the notion that the juxtaposition of a variety of processing methods provides improved retrieval performance over and above the performance of the individual component methods. The design philosophy of the SMART system, which is based on an iterative search procedure with a variety of analysis methods to retrieve relevant information, should therefore prove useful in practice. (A similar conclusion, pointing to the joint use of UDC (Universal Decimal Classification) coupled to a Uniterm system, has previously been reached in a conventional retrieval situation.[11])

Additional experiments remain to be carried out with different document collections not previously used with the available dictionaries, and with additional search requests. A careful analysis of systems failures is also mandatory, in order to determine more precisely the

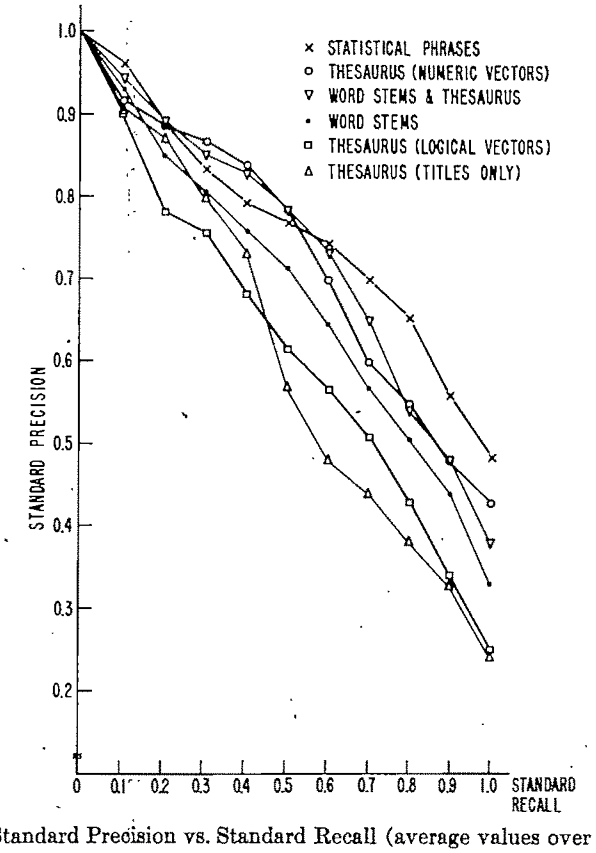


FIG. 11. Standard Precision vs. Standard Recall (average values over 17 requests).

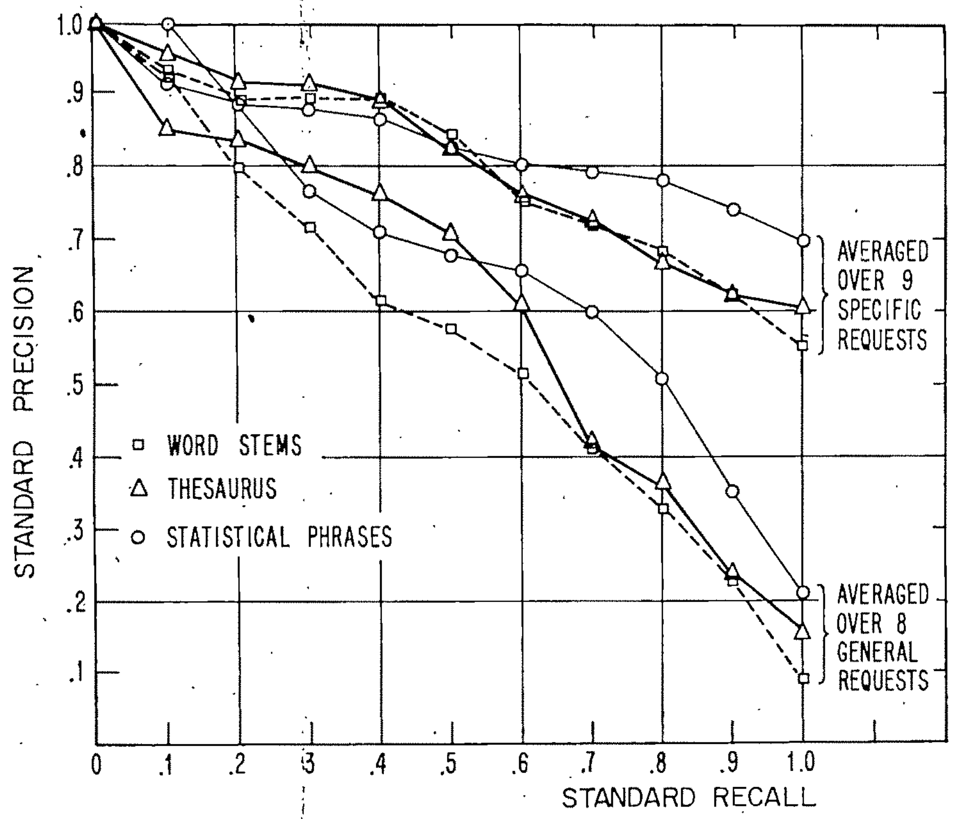


FIG. 12. Standard Precision vs. Standard Recall
Comparison of General and Specific Requests.

strengths and weaknesses of the individual methods, and the circumstances under which relevant documents are not recognized and receive therefore a low correlation on the output lists. Additional processing sequences must also be analyzed, and useful sequences identified in order to maximize system performance and retrieval effectiveness.

References

1. G. SALTON, "A Document Retrieval System for Man-machine Interaction," *Proceedings of the 19th ACM Annual Conference*, Philadelphia, 1964.
2. ——— and M. E. LESK, "The SMART Automatic Document Retrieval System—An Illustration," *Communications of the ACM*, vol. 8, no. 6, June 1964.
3. ——— et al., "Information Storage and Retrieval," Reports No. ISR-7 and ISR-8 to the National Science Foundation, The Computation Laboratory of Harvard University, June 1964 and December 1964.
4. J. A. SWETS, *Information Retrieval Systems Science*, vol. 141, no. 3577, July 19, 1963.
5. C. W. CLEVERDON, "The Testing of Index Language Devices," *ASLIB Proceedings*, vol. 15, no. 4, April 1963.
6. ———, "The Testing and Evaluation of the Operating Efficiency of the Intellectual Stages of Information Retrieval Systems," International Study Conference on Classification Research, Elsinore, September 1964.
7. D. R. SWANSON, "Searching Natural Text by Computer," *Science*, vol. 132, no. 3434, October 21, 1960.
8. ———, "The Evidence Underlying the Cranfield Results," *The Library Quarterly*, vol. 35, no. 1, January 1965.
9. J. ROCCHIO, "Performance Indices for Document Retrieval Systems," "Information Storage and Retrieval," Report No. ISR-8 to the National Science Foundation, The Computation Laboratory of Harvard University, December 1964.
10. R. A. FAIRTHORNE, "Basic Parameters of Retrieval Tests," short paper presented at the 1964 ADI Annual Meeting, Philadelphia, October 1964.
11. J. A. SCHULLER, "Experience with Indexing and Retrieving by UDC and Uniterms," *ASLIB Proceedings*, vol. 12, no. 11, November 1960.

Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing*

A detailed comparison of how 334 papers in Vol. 112 of *Physical Review* form related groups according to two criteria of relatedness. The criteria are: (1) the Analytic Subject Index as used by the editors of the *Physical Review;* and (2) the method of Bibliographic Coupling. The similarities and differences between the groups thus formed are illustrated and discussed.

M. M. KESSLER

The Libraries, Massachusetts Institute of Technology
Cambridge, Massachusetts

Introduction

Previous reports in this series[1-8] described and illustrated a new method for separating a large body of technical literature into small related groups. This method was called "Bibliographic Coupling" because it originated in the hypothesis that the bibliography of technical papers is one way by which the author can indicate the intellectual environment within which he operates, and if two papers show similar bibliographies, there is an implied relation between them. The previous reports demonstrated that the phenomenon of bibliographic coupling exists and that the titles so coupled do indeed show a relatedness in subject matter.†

The question naturally arises, how does this method of grouping papers compare with the results of certain standard and well-known techniques currently in use? Since the bulk of our experimental material derives from physics literature and, more specifically, from the *Physical Review,* we shall perform the present comparison also on a sample volume of the *Physical Review.*

It is clear that any comparison of the results of two complex processes will reflect the subjective bias and criteria of the author. It is not just a matter of displaying the author's favoritism or possessive pride. Such crude prejudices can be overcome, and if not they are easily detected by the reader. But the mental bias and the perfectly legitimate intellectual convictions of the author cannot, and perhaps should not, be eliminated. The subjective bias is stamped on the problem and on its solution by the choice of questions that the research is supposed to answer.

The question—How do two processes compare with regard to their ability to form groups of technical papers?—and the question—How do two processes compare with regard to solving the problems of technical communication?—are two different orders of question from the logical point of view. The first question can be answered entirely within the domain of the two processes and their results. Indeed, one may be purely formal and assign arbitrary numbers and letters to the papers and to the subject categories and make the comparison without any reference to the titles and subject fields involved. The result of such a study is apt to be a set of numbers and statistics of considerable precision. But this precision is *not* transferrable to the second order of questions. There we must go out of the domain of the two processes and seek judgment that is relevant to human needs. But such needs reflect complex intellectual habits, tastes and shades of meaning. Do what we will with our statistics, graphs and tables, we must at last face the fact that we are dealing with a field of experience where there is no single criterion that can be applied or definition sharply drawn. A group of papers is more or less

* Supported by a grant from the National Science Foundation to the M.I.T. Libraries and in part by Project MAC, M.I.T. research program sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research.

† The essentials of the method of Bibliographic Coupling are:

a. A single item of reference used by two papers is called one *unit of coupling* between them.
b. A number of papers constitute a related group, G_A, if each member of the group has at least one coupling unit to a *given test paper* P_0.
c. The *coupling strength* between P_0 and any member of G_A is measured by the number of coupling units (n) between them.

what a reader has asked for. A bibliographic list is more or less complete. A given paper fits more or less in a subject field category. The pretension under such circumstances to be rigorously "scientific" and numerically "exact" would only mislead us in our judgment.

Tests and comparisons designed to answer questions of the first order can profitably be performed within the limited environment of the laboratory. Questions of the second order, however, are more suitably tested in a carefully controlled but realistically functional environment that simulates many of the influences and pressures of the real world. It is only to the first type of question that we address ourselves in this report.

Our experimental material consisted of 334 papers in Vol. 112 (1958) of the *Physical Review.* These papers were arranged into groups according to two schemes: the Analytic Subject Index (ASI) and Bibliographic Coupling (BC). In the first instance, the editors supply a list of 73 subject categories. Someone in the editorial office considers each paper separately and places it into one or more categories. At the end of this process, each of the 334 papers finds itself in a subject category. Groups of papers are thus formed under the heading of each subject category. There will be as many groups as there are categories, in this case 73, although some of the groups may be empty. In the second case (BC) each paper is compared to each of the remaining 333. Those that share one or more bibliographic items with the test paper are considered to be members of a group. There will be as many groups as there are papers in the sample, in this case 334, although some of the groups may be empty.

Table 1 shows the group-forming characteristics of the two methods.

Although there is no upper limit to the number of categories in which a given paper may be placed, we find that the practitioners of ASI in the *Physical Review* try to keep this number well below four. Of the 334 papers, 159 are in one category only, 124 find themselves in two groups, 47 in three groups, and only four papers are members of four groups. In the case of BC, the distribution of papers among groups is the same as the distribution of groups having N papers (third column in Table 1); 44 papers fall into no group at all, 56 in one group only, 45 in two groups, 30 in three groups, etc. We see that in ASI a collection of papers will be distributed among a smaller number of groups and each paper will be a member of fewer groups than in BC. This is of no particular significance at this time since both methods generate enough groups to allow comparison. We must, however, point out the following significant differences in the mechanics of group formation because they will have to be appealed to later when we examine the results of the two methods.

In the case of BC, the process for group formation follows a fixed prescription. Even when performed by a human being, it is completely mechanical. There is no bending or twisting to make a fit; no discretion or leeway whatsoever is available to the operator. The coupling criterion is fixed and the results were determined once and for all by the authors who wrote the paper. Not so in the case of ASI. Here much is left to the discretion of the editors and indexers. The very choice of categories reflects the expectation of the editors. In areas where large numbers of papers are expected, a much finer subdivision of field categories is provided than in areas where the expectation of papers is low. Thus we have a class "Liquid Helium" referring to a particular state of one element and a class "Nuclear Reactions Induced by Neutrons" equally narrow in scope, while at the same time we have "Biophysics," "Geophysics," and "Acoustics," which cover major scientific fields. This redeflects the editors' expectation that there will be very few papers in Biophysics, Geophysics, and Acoustics in the *Physical Review.* Or consider the class "Nuclear Reactions—General" in addition to five classes for nuclear reactions each induced by a specific particle. Clearly the editors take a pragmatic view of the field and divide it into 73 cells of equal expectation rather than 73 logical subdivisions. This leads to a reasonable distribution of papers among cells with a minimum of empty categories. It also leads to a lack of uniformity in the relatedness of two papers that find themselves in one cell. For example, two papers assigned to the category (or cell) Liquid Helium would be related much more closely to each other than two papers assigned to the category Astrophysics.

Table 1. Groups formed according to ASI and BC.

Number of papers in the group (N)	Number of groups having N papers ASI	BC
0	10	44
1	9	56
2	5	45
3	6	30
4	7	31
5	5	17
6	3	18
7	2	11
8	5	6
9	5	14
10	5	13
11	0	10
12	0	10
13	2	6
14	0	4
15	2	6
16	2	5
17	1	2
18	0	1
20	1	1
21	1	1
23	1	0
24	0	1
25	0	1
26	0	1
27	1	0
38	1	0
39	1	0
49	1	0

In the case of BC, all groups should exhibit the same degree of relatedness. But since this relatedness is statistical, there is in actual practice a distribution of relatedness *within* each group, although the properties from group to group are more or less uniform.

Before proceeding with the actual comparison of the groups formed by the two methods, we remind the reader that in this test we are concerned merely with the mechanics of group formation and not with its meaning or relevance to retrieval. If we find an equivalence between the two sets of groups, they may be equally good, equally bad or equally indifferent as far as a given mode of retrieval is concerned. In other words, we address ourselves to questions of the first order. We do this in order to delineate clearly the problem of this report and to avoid confusion. We fully recognize that questions of the second order are of prime significance. But they must be postponed. Since, however, all the papers in our illustrations are in Vol. 112 of the *Physical Review* and are identified by page number, the reader is free to refer to them if he so desires and form his own judgment as to their relevance to some retrieval application that he may have in mind.

The comparison will be performed in four stages:

1. Given the papers in a group according to the BC criterion, how are they regarded by the ASI criterion?

2. Given the papers in a group according to the ASI criterion, how are they regarded by the BC criterion?

3. Given two papers considered to be strongly related by BC, what is the verdict of ASI?

4. Given two papers considered to be strongly related by ASI, what is the verdict of BC?

• First Stage

Each of the 334 papers in Vol. 112 was used as a test paper to generate a group according to BC. For example, the paper on page 624 acting as P_0 generated a group of fourteen papers as shown in the first row in Table 2 under G_A. We then consulted Vol. 112 of the *Physical Review* to see how each of these papers was classified according to ASI. This is indicated by an X in the appropriate row. Thus we see that paper 273 was placed in three ASI categories, paper 614 in two, etc. For the group as a whole, we find that, according to the ASI criterion, all but one belonged together in the category headed "Elementary Particle Interactions." Ten of the fourteen papers were judged to belong in the category "Mesons and Hyperons." Although seven of the available 73 ASI categories were used to describe the fourteen papers in this G_A, two of these categories would have accounted for the entire list with some redundancy to spare. This G_A shows a particularly high homogeneity with respect to the judgment of ASI. At the other extreme, however, consider the group generated by $P_0 = 968$ as shown in Table 3. There are nine papers in this group and, in order to account for all of them, we need six of the eight ASI categories used. No single ASI category contains more than two of the nine papers and there is very little redundancy. A formal appraisal of Table 3 would suggest a low correlation between the results of BC and ASI. However, an examination of the ASI categories shows that this only slightly correlated case is not

Table 2. First stage comparison of BC and ASI (P_o=624; see text for explanation).

ASI Categories G_A	273	605	614	642	979	981	986	1303	1311	1793	1813	2053	2107	2135
Elementary particle interactions	X	—	X	X	X	X	X	X	X	X	X	X	X	X
Mesons and hyperons	X	X	X	—	X	X	—	—	X	X	—	X	X	X
Field theory	X	—	—	X	—	—	—	—	—	—	X	—	—	X
Nuclear photoeffects	—	—	—	—	—	X	—	—	—	—	—	X	—	—
Scattering of mesons and hyperons	—	—	—	—	—	—	X	—	—	—	X	—	—	—
Scattering of protons	—	—	—	—	—	—	—	X	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	—	—	X	—	—	—	—	—

Table 3. First stage comparison of BC and ASI (P_o=968; see text for explanation).

ASI Categories G_A	232	580	906	1295	1325	1658	1667	1719	1810
Mesons and hyperons	—	X	—	—	—	—	—	—	X
Scattering of mesons and hyperons	—	—	—	—	X	—	—	—	—
Nuclear reactions induced by deuterons and tritons	—	—	—	—	—	—	X	—	—
Scattering of protons	—	—	—	—	—	X	—	—	—
Nuclear spectra	X	—	—	—	—	—	—	—	—
Nuclear structure theory	—	—	X	—	—	—	—	X	—
Mathematical methods	—	—	X	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	X	—	X	—	—	—

as bad as it looks because the ASI categories are logically very close to each other, and a slight change in emphasis by the indexer could have changed the picture materially. Similar tables were made for each of the 334 groups generated by the method of BC. To keep this report to manageable size, we shall not reproduce all the data. We give a sizeable sample of the results and a statistical summary of the whole. Table 4 shows the comparison of five groups, each consisting of ten papers. Table 5 shows the results of five groups each containing fifteen papers. Table 6 shows five groups of 20 or more papers. Table 7 is a statistical summary of the results of all of the 334 papers.

To obtain an empirical measure of how the results of the two methods compare, we consider that if all the members of G_A could be accounted for by three ASI categories there is a good match between the two methods. When that is not so we use the additional measure of the fraction of ASI categories necessary to account for all the members of the group.

The summary data (Table 7) are given in the following way. Each paper has four numbers associated with it. These numbers are given as two fractions $^A/_B$ and $^C/_D$ where:

A = the largest number of papers in G_A included in three ASI categories.

B = the number of papers in G_A.

C = the number of ASI categories that would account for all the papers in G_A.

D = the total number of ASI categories used to describe all the papers in G_A.

A little consideration will show that a "good" correlation between the results of BC and ASI will be characterized by a relatively large value of $^A/_B$ and a small value of $^C/_D$. The ratio of the two fractions may be used as a measure of the correlation. Thus in the example of Table 2 this ratio is $\frac{14}{[illegible]}/\frac{2}{7}=3.5$; in the example of Table 3 the ratio is $\frac{8}{9}/\frac{8}{8}=.9$. We do not attach any theoretical significance to this ratio. It is used here as a matter of convenience.

In evaluating the results of this comparison as summarized in the table above, we eliminated papers that formed BC groups of three or less. This was necessary because we happened to pick three ASI categories as a reasonable measure of agreement between the results of BC and ASI. The average value of $^A/_B$ for all BC groups having four papers or more is 0.9. This means that, on the average, the a priori verdict of ASI indexing is that three ASI categories will account for nine out of every ten papers in the BC groups. Had we not eliminated the smaller groups this number would have been even larger. For the same group of papers the average value of the fraction $^C/_D$ is 0.5. This means that one-half of all the ASI categories used by the editors to index the papers in each BC group would have sufficed to account for all the members of the group.

• Second Stage

In this experiment, we start with the papers found under a given ASI heading and call them a group. We then observe the distribution of the members of this group among the BC groups. It will be recalled that the nature of BC is such that every paper generates a group. If, therefore, we find that a certain ASI heading contains 15 papers, we expect each of these to generate a BC group. The papers in the ASI group will thus be distributed among 15 BC groups. The question that we must answer is—Are these groups more alike than one might expect from random association? A measure of this answer might be to note how many of the 15 BC groups are necessary to account for all of the 15 papers in the single ASI group. An example will help clarify what we mean. Consider the group of papers under the ASI heading "Elementary Particle Interactions." There are 36 papers in this group. These are shown in the first column of Table 8. Each of these papers generates a group by the BC method. Thus paper 624 generates a group that includes, among others, 13 of the 36 papers in column 1. These are marked X in column 2. Paper 986 generates a BC group that includes 12 of the 36 papers. These are marked X in column 3. These two BC groups account for 19 of the 36 papers in the given ASI category. If we continue this process we find that 8 BC groups account for 33 of the 36 papers. The remaining papers (587, 1335 and 1642) do not form groups according to BC. This is a highly correlated example. A less correlated example is given by the ASI category "Crystalline State," Table 9. There are 45 papers in this category. The BC groups generated by seven of these account for 21. Beyond that no single BC group contributes more than one paper to the category. The data in general fall between these two extremes, depending on the "logical size" of the ASI categories.

As was pointed out previously in this report, the ASI categories were not chosen entirely on the basis of a logical subdivision of the subject field. Allowances were made for the expected number of papers in each category. The amount of relatedness that one may expect within each ASI category varies. It is therefore to be expected that when we apply a fixed measure of relatedness to the papers in each category, the results will show extreme variations. Some of the tight categories like "Gamma Rays" or "Properties of Films" will gather papers highly related and these will show great correlation when compared to the results of BC. On the other hand, categories like "Geophysics," "Mechanics" and "Biophysics," to say nothing of "Miscellaneous," will gather a few papers clearly to be included in the subject categories, but not necessarily related.

Table 4. First stage comparison of BC and ASI: five groups of ten papers each.

(a) $P_o = 344$

ASI Categories \ G_A	31	90	186	322	334	812	935	1092	1137	1169
Crystalline state	—	X	—	X	X	X	—	X	X	—
Superconductivity	X	—	—	—	—	—	—	—	—	—
Dielectrics and dielectrical properties	—	X	—	—	—	—	—	—	—	—
Atomic and molecular beams	—	—	X	—	—	—	—	—	—	—
Magnetic resonance	—	—	X	—	—	—	X	—	—	X
Nuclear moments and spin	—	—	X	—	—	—	X	—	—	—
Elasticity and plasticity	—	—	—	—	—	—	—	X	—	—
Spectra, general	—	—	—	—	—	—	—	—	—	X

(b) $P_o = 1227$

ASI Categories \ G_A	503	547	897	931	1224	1658	1975	1989	2020	2029
Nuclear reactions, general	—	—	—	—	—	—	—	—	—	X
Nuclear reactions induced by deuterons and tritons	X	—	—	—	X	—	X	—	X	—
Nuclear reactions induced by α particles and He^3	—	X	—	X	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	X	—	—	X	—	—	—	—
Nuclear spectra	—	—	—	X	—	—	—	—	—	—
Scattering of protons	—	—	—	—	—	X	—	—	—	—
Scattering of α particles	—	—	—	—	—	—	—	X	—	—
Nuclear structure theory	—	—	—	—	—	—	—	—	—	X
Scattering, general	—	—	—	—	—	—	—	—	—	X

(c) $P_o = 1311$

ASI Categories \ G_A	271	614	624	906	1027	1287	1295	1303	1319	1802
Elementary particles	—	X	X	—	X	X	—	X	—	X
Mesons and hyperons	—	X	X	—	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	X	—	X	—
Scattering, general	X	—	—	—	—	—	—	—	—	—
Mathematical methods	—	—	—	X	—	—	—	—	—	—
Nuclear structure theory	—	—	—	X	—	—	—	—	—	—
Field theory	—	—	—	—	X	—	—	—	—	X
Scattering of mesons and hyperons	—	—	—	—	—	X	—	—	—	X
Scattering of protons	—	—	—	—	—	—	—	X	—	—

(d) $P_o = 1900$

ASI Categories \ G_A	31	328	812	906	994	1008	1056	1083	1522	1888
Superconductivity	X	—	—	—	—	—	—	X	X	X
Crystalline state	—	X	X	—	—	—	—	—	—	—
Gases	—	X	—	—	—	—	—	—	—	—
Magnetic properties	—	X	—	—	—	—	—	—	—	—
Mathematical methods	—	—	—	X	—	—	—	—	—	—
Nuclear structure theory	—	—	—	X	X	X	—	—	—	—
Field theory	—	—	—	—	X	X	—	—	—	—
Statistical mechanics	—	—	—	—	X	X	X	—	—	—
Thermal properties	—	—	—	—	—	—	—	X	—	—

(e) $P_o = 2020$

ASI Categories \ G_A	503	890	903	923	1227	1658	1667	1975	2029	2048
Nuclear reactions induced by deuterons and tritons	X	—	—	X	—	—	X	X	—	—
Nuclear photoeffects	—	X	—	—	—	—	—	—	—	—
Nuclear spectra	—	—	X	—	—	—	—	—	—	—
Nuclear structure theory	—	—	—	X	—	—	—	—	X	—
Nuclear reactions, general	—	—	—	—	X	—	—	—	X	—
Nuclear reactions induced by protons	—	—	—	—	—	X	—	—	—	—
Scattering of protons	—	—	—	—	—	X	—	—	—	—
Scattering, general	—	—	—	—	—	—	—	—	X	—
Scattering of α particles	—	—	—	—	—	—	—	—	—	X

Table 5. First stage comparison of BC and ASI: five groups of fifteen papers each.

(a) $P_o = 503$

ASI Categories \ G_A	232	452	461	903	923	931	1200	1227	1252	1337	1375	1763	1975	2020	2029
Nuclear reactions induced by deuterons and tritons	—	—	X	—	X	—	—	—	—	—	—	—	X	X	—
Nuclear spectra	X	—	—	X	—	X	—	—	X	—	—	—	—	—	—
Nuclear structure theory	—	X	—	—	X	—	—	—	—	—	—	—	—	—	X
Nuclear reactions induced by α particles and He^3	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	X	—	X	—	—	—	—	—	—
Nuclear reactions, general	—	—	—	—	—	—	—	X	—	—	—	—	—	—	X
Scattering of mesons and hyperons	—	—	—	—	—	—	—	—	—	X	—	X	—	—	—
Elementary particles	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Field theory	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Scattering, general	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X

(b) $P_o = 554$

ASI Categories \ G_A	461	468	518	560	568	897	954	963	1183	1267	1295	1679	1730	1994	2026
Nuclear photoeffects	—	—	—	X	—	—	—	—	—	—	—	—	—	X	—
Nuclear reactions induced by deuterons and tritons	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear spectra	X	X	—	—	—	—	—	—	—	—	—	—	X	—	X
Scattering of electrons and positrons	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	X	—	—	—	—	X	X	—	—	—
Nuclear structure theory	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—
Elementary particles	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Neutrinos and antineutrinos	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Atomic structure and spectra	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—
X–rays	—	—	—	—	—	—	—	—	X	X	—	—	—	—	—
Radiation	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

(c) $P_o = 897$

ASI Categories \ G_A	518	532	554	560	935	1183	1227	1247	1658	1672	1702	1730	1958	2026	2029
Nuclear reactions induced by protons	—	X	—	—	—	—	—	—	X	X	X	—	—	—	—
Nuclear spectra	X	—	—	—	—	—	—	X	—	—	—	X	X	X	—
X-rays	—	X	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear photoeffects	—	—	X	X	—	—	—	—	—	—	—	—	—	—	—
Magnetic resonance	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Nuclear moment and spin	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Atomic structure and spectra	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear reactions, general	—	—	—	—	—	—	X	—	—	—	—	—	—	—	X
Scattering of protons	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—
Scattering, general	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X

(d) $P_o = 1027$

ASI Categories \ G_A	614	669	866	986	994	1008	1287	1311	1344	1370	1375	1380	1555	1813	1843
Elementary particles	X	—	—	X	—	—	X	X	X	—	X	X	—	X	—
Field theory	—	X	—	—	X	X	—	—	X	—	X	X	—	X	X
Mesons and hyperons	X	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Magnetic resonance	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—
Molecular structure and spectra	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—
Scattering of mesons and hyperons	—	—	—	X	—	—	X	—	—	X	—	—	—	X	—
Nuclear structure theory	—	—	—	—	X	X	—	—	—	—	—	—	—	—	—
Statistical mechanics	—	—	—	—	X	X	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Scattering, general	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Crystalline state	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Dielectrics and dielectrical properties	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Quantum mechanics	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
Radiation	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X

Table 5—Continued

(e) $P_o = 2135$

ASI Categories G_A	267	273	580	605	624	642	665	979	981	986	1375	1746	1793	2053	2128
Elementary particles	X	X	—	—	X	X	X	X	X	X	X	—	X	X	X
Field theory	—	X	—	—	—	X	—	—	—	—	X	—	—	—	—
Mesons and hyperons	X	X	X	X	X	—	X	X	X	—	—	X	X	X	X
Quantum electrodynamics	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear photoeffects	—	—	—	—	—	—	—	—	X	—	—	—	—	X	—
Scattering of mesons and hyperons	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—

Table 6. First stage comparison of BC and ASI: five groups of twenty or more papers each.

(a) $P_o = 232$

ASI Categories G_A	244	267	273	452	461	503	580	605	665	877	893	968	979	1240	1252	1337	1375	1702	1746	1763	1975	2029	2061	2128	2135
Elementary particles	X	X	X	—	—	—	—	—	X	—	—	—	X	—	—	—	X	—	—	—	—	—	X	X	X
Neutrinos and anti-neutrinos	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear spectra	X	—	—	—	—	—	—	—	—	X	—	—	—	X	X	—	—	—	—	—	—	—	—	—	—
Mesons and hyperons	—	X	X	—	—	—	X	X	X	—	X	X	X	—	—	—	—	—	X	—	—	—	X	X	X
Quantum electrodynamics	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Field theory	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	X
Nuclear structure theory	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Nuclear reactions induced by deuterons and tritons	—	—	—	—	X	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Radioactivity	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Scattering of protons	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by mesons and hyperons	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	X	—	—	—	—	—	—	—
Scattering of mesons and hyperons	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	X	—	—	—	—	—
Nuclear reactions, general	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Scattering, general	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—

(b) $P_o = 273$

ASI Categories G_A	232	244	481	580	605	614	624	665	877	893	945	963	979	1240	1375	1637	1684	1998	2061	2135
Nuclear spectra	X	X	X	—	—	X	X	X	X	X	X	—	—	X	—	—	X	X	—	—
Elementary particles	—	X	—	—	—	—	—	—	—	—	—	X	X	—	X	X	—	—	X	X
Neutrinos and anti-neutrinos	—	X	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—
Mesons and hyperons	—	—	—	X	X	X	X	X	—	—	—	—	X	—	—	—	—	—	X	X
Radioactivity	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—
Scattering of protons	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Field theory	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	X
Atomic structure and spectra	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Electrons and positrons	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Quantum electrodynamics	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—

(c) $P_o = 605$

ASI Categories G_A	232	267	273	580	614	624	665	979	986	1021	1282	1335	1375	1658	1719	1746	1773	2077	2118	2128	2135
Mesons and hyperons	—	X	X	X	X	X	X	X	—	—	—	X	—	—	—	X	—	—	X	X	X
Nuclear spectra	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Elementary particles	—	X	X	—	X	X	X	X	X	X	—	X	X	—	—	—	—	X	—	X	X
Quantum electrodynamics	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Field theory	—	—	X	—	—	—	—	—	—	X	—	—	X	—	—	—	—	—	—	—	X
Scattering of mesons and hyperons	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	X	—	—	—	—	—
Scattering of electrons and positrons	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	X	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Scattering of protons	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Nuclear structure theory	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—

Table 6—Continued

(d) $P_o = 1375$

ASI Categories	G_A	232	267	273	452	461	503	580	605	665	931	945	979	1027	1252	1337	1637	1746	1763	1796	1813	1975	2010	2029	2128	2135
Elementary particles		—	X	X	—	—	—	—	—	X	—	—	X	X	—	—	X	—	—	X	X	—	—	—	X	X
Field theory		—	—	X	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	X	—	—	—	—	X
Nuclear spectra		X	—	—	—	—	—	—	—	—	X	X	—	—	X	—	—	—	—	—	—	—	X	—	—	—
Mesons and hyperons		—	X	X	—	—	—	X	X	X	—	—	X	—	—	—	—	X	—	X	—	—	—	—	X	X
Quantum electrodynamics		—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear structure theory		—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Nuclear reactions induced by deuterons and tritons		—	—	—	—	X	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Nuclear reactions induced by α particles and He^3		—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	X	—	X	—	—	—	—	—
Atomic structure and spectra		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Electrons and positrons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear reactions, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—

(e) $P_o = 2029$

ASI Categories	G_A	186	232	452	461	503	518	855	897	903	923	931	986	1192	1227	1252	1337	1375	1658	1672	1763	1958	1975	1981	1989	2020	2048
Atomic and molecular beams		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Magnetic resonance		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear moment and spin		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear spectra		—	X	—	—	—	X	—	—	X	—	—	—	—	—	X	—	—	—	—	—	X	—	—	—	—	—
Nuclear structure theory		—	—	X	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by deuterons and tritons		—	—	—	X	X	—	—	—	—	X	X	—	—	—	—	—	—	—	—	—	—	X	—	—	X	—
Atomic structure and spectra		—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Spectra, general		—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	X	—	—	—	—	—	—	X	—	—	X	X	—	—	—	—	—	—	—
Elementary particles		—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	X	—	—	—	X	—	—	—	—	—	—
Nuclear reactions induced by neutrons		—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—
Scattering of neutrons		—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions, general		—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—
Field theory		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Scattering of protons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—	—
cattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Scattering of α particles		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	X	—	X

Table 7. Statistical summary of first stage comparison. Page numbers refer to title page of articles in Vol. 112. See text for explanation of the variables A/B and C/D. See also remarks following this table.

Page	A/B	C/D	Page	A/B	C/D
31	7/7	3/6	369	4/4	1/[illegible]
49	4/4	3/6	381	4/4	1/6
80	4/4	2/7	397	4/4	1/4
136	5/6	3/5	452	9/18	6/11
179	5/5	3/7	461	11/17	8/13
182	4/4	2/4	468	8/10	5/6
186	12/13	4/9	503	10/15	7/10
191	8/9	4/6	512	4/4	2/4
203	4/5	4/5	532	10/12	5/7
229	4/4	1/2	547	6/7	4/6
262	6/8	3/5	554	9/15	8/11
267	11/12	4/9	560	8/11	6/8
271	5/5	3/9	580	16/18	6/11
298	5/6	4/5	605	17/21	7/10
317	6/6	3/6	614	9/9	2/5
322	5/5	1/2	618	6/6	2/7
326	4/4	1/4	624	14/14	2/7
328	4/5	4/5	642	12/12	3/5
334	4/4	1/1	665	12/12	3/7
337	4/4	3/4	685	8/8	3/7
344	10/10	3/8	698	4/4	2/5

Table 7—Continued

Page	A/B	C/D	Page	A/B	C/D
715	5/5	2/6	1504	4/4	2/8
732	6/6	2/6	1522	6/7	4/5
812	11/16	7/16	1599	4/4	1/4
828	4/4	3/6	1603	5/5	1/5
843	5/6	4/7	1620	4/4	1/3
855	4/4	3/8	1637	5/5	2/7
870	4/4	3/6	1658	8/16	9/13
890	4/4	2/3	1667	5/6	4/7
897	11/15	6/10	1672	7/9	5/6
906	11/15	7/12	1719	5/10	8/13
923	5/6	4/6	1730	8/9	4/5
931	5/7	5/9	1746	12/12	3/5
945	12/12	3/7	1759	4/4	3/7
954	5/7	5/6	1763	9/14	7/12
963	7/8	4/7	1773	9/11	6/8
968	6/10	7/10	1793	8/8	3/6
979	12/12	3/5	1802	5/5	3/7
981	7/7	1/5	1813	13/13	3/6
986	14/16	5/11	1877	6/6	3/5
992	4/4	1/4	1888	6/6	3/4
994	9/15	8/15	1900	9/10	4/10
1008	6/10	7/13	1950	5/5	2/5
1021	10/13	6/9	1954	10/10	3/6
1027	13/15	5/14	232	19/25	7/15
1056	5/6	4/8	273	20/20	3/10
1083	5/5	4/5	1975	8/17	10/13
1092	6/6	3/5	1981	4/4	2/4
1112	4/4	2/3	1989	6/7	4/7
1130	4/4	1/4	1994	5/5	3/4
1151	4/4	2/4	2010	4/4	3/4
1157	5/5	3/6	2020	7/10	6/9
1183	7/9	5/8	2043	5/6	4/6
1192	9/12	6/10	2048	5/9	7/10
1200	9/10	4/7	2053	9/9	2/9
1210	4/5	4/6	2072	5/5	3/8
1217	5/7	5/6	2077	10/14	7/11
1227	8/10	5/9	2112	4/4	2/3
1247	7/7	3/6	2128	9/9	3/5
1252	8/16	8/14	2135	6/6	7/7
1257	8/8	3/5	2026	9/11	5/9
1267	5/	2/6	2029	14/26	10/18
1282	6/9	6/10	1169	6/6	2/3
1287	5/8	3/7	1627	9/10	4/13
1295	9/13	7/12	1958	8/12	6/12
1303	9/9	8/10	518	9/14	6/12
1311	9/10	4/9	481	12/12	3/10
1319	7/7	3/9	489	4/4	2/5
1337	8/11	5/9	903	7/10	6/10
Page	A/B	C/D	935	8/9	4/8
1344	9/10	4/7	877	9/9	2/6
1367	4/4	2/4	893	11/11	2/6
1370	7/7	2/3	1240	11/11	3/9
1375	18/25	6/14	1684	10/10	3/6
1391	6/7	4/13	1702	8/11	5/8
1404	5/5	2/5	1998	11/11	2/8
1429	6/6	3/5	2004	8/8	3/5
1456	4/5	4/9	2061	12/13	4/8
1488	4/4	2/7	481	12/12	3/10

In order to analyze the results of this comparison statistically, we must have a measure of relatedness for the ASI categories. This would lead us into a measure of ASI independent of its comparison to BC, a task that we do not undertake at this time. Of course, one could claim that the extent of correlation with the results of BC is itself a measure of the "size" of the ASI category. Such an attitude, however, cannot properly be maintained in the course of an experiment that involves BC as a variable. We, therefore, draw the qualitative conclusion that in those cases where the ASI categories seem to be of a small enough "logical size," the correlation with BC is good. When the "logical size" of the categories broadens to a point where sizeable fields of knowledge are involved, the correlation diminishes.

• Third Stage

At this time it will profit the reader to review the defining statement on Bibliographic Coupling as given in

Table 8. Second stage comparison of BC and ASI.
Column A=papers in the category "Elementary Particle Interactions."
Subsequent columns are the BC groups generated by the papers indicated at the top of the column.
See text for explanation and comments.

A	624	986	1813	2135	273	642	1344	981
244	—	—	—	—	X	—	—	—
267	—	—	—	X	—	—	—	—
273	X	—	—	X	—	—	—	—
283	—	—	—	—	—	X	—	—
587	—	—	—	—	—	—	—	—
614	X	—	—	—	X	—	—	—
618	—	—	X	—	—	—	—	—
622	—	—	—	—	—	—	—	X
624	—	X	X	X	X	X	—	X
642	X	X	X	X	—	—	X	X
665	—	—	—	X	X	—	—	—
963	—	—	—	—	X	—	—	—
979	X	—	—	X	X	—	—	—
981	X	X	—	X	—	X	—	—
986	X	—	X	X	—	X	X	X
992	—	—	—	—	—	—	X	—
1021	—	X	X	—	—	X	—	—
1027	—	X	X	—	—	—	X	—
1287	—	X	X	—	—	—	X	—
1303	X	—	X	—	—	—	—	—
1311	X	—	—	—	—	—	—	—
1335	—	—	—	—	—	—	—	—
1344	—	X	X	—	—	X	—	—
1375	—	—	X	X	X	—	—	—
1380	—	—	—	—	—	—	X	—
1637	—	—	—	—	X	—	—	—
1642	—	—	—	—	—	—	—	—
1793	X	X	—	X	—	X	—	X
1802	—	—	X	—	—	—	—	—
1813	X	X	—	—	—	X	X	—
2053	X	X	—	X	—	X	—	X
2061	—	—	—	—	X	—	—	—
2077	—	X	X	—	—	—	—	—
2107	X	—	—	—	—	—	—	—
2128	—	—	—	X	—	—	—	—
2135	X	X	—	—	X	X	—	X

the footnote on the first page. The statement indicates two conditions of relatedness between papers. A group of papers may be related by virtue of each being a member of a given G_A, but also each member of G_A forms a pair relation with P_0. The first instance produces groups of papers; the second gives pairs of papers. The first and second stages of comparison considered how the various G_A groups compare with groups of papers formed by assignment to ASI categories. In the present comparison, we consider the pairs of papers formed between P_0 and each member of G_A.

The number of pairs formed is equal to one-half of the cumulative total of the papers in all the G_A groups. In Vol. 112 of the *Physical Review,* 876 pairs of papers were formed variously coupled according to the indicated coupling strength shown in Table 10.

Since there are very few pairs of high coupling strength we place all those with n = 5 and larger into one group.

* The random pairs were chosen by picking 200 telephone numbers, two at a time, from a random place in the directory. The nearest page numbers corresponding to the telephone numbers were called a random pair of papers.

Table 9. Second stage comparison of BC and ASI.
Column A=papers in the category "Crystalline State."
Subsequent columns are the BC groups generated by the papers indicated at the top of the column.
See text for explanation and comments.

A	344	715	1092	1603	322	381	1620
54	—	—	—	—	—	—	—
64	—	—	—	—	—	—	—
73	—	—	—	—	X	—	—
80	—	X	—	—	—	X	—
90	X	—	—	—	—	—	—
103	—	—	—	—	—	—	—
322	X	—	X	—	—	—	—
328	—	—	—	—	—	—	—
334	X	—	—	—	X	—	—
337	—	—	—	—	—	—	X
344	—	—	X	—	X	—	—
381	—	X	—	—	X	—	—
685	—	—	—	—	—	—	—
698	—	X	—	—	—	—	—
708	—	X	—	—	—	—	—
715	—	—	—	—	—	X	—
722	—	—	—	—	—	—	X
725	—	—	—	—	—	—	—
751	—	—	—	—	—	—	—
765	—	—	—	—	—	—	—
785	—	—	—	—	—	—	—
804	—	—	—	—	—	—	—
812	X	—	—	—	—	—	—
837	—	—	—	—	—	—	—
1058	—	—	—	—	—	—	—
1061	—	—	—	—	—	—	—
1063	—	—	—	—	—	—	—
1092	X	—	—	—	X	—	—
1117	—	—	—	—	—	—	—
1130	—	—	—	X	—	—	—
1137	X	—	—	—	—	—	—
1516	—	—	—	X	—	X	—
1534	—	—	—	—	—	—	—
1540	—	—	—	—	—	—	—
1555	—	—	—	—	—	—	—
1567	—	—	X	—	—	—	—
1571	—	—	—	—	—	—	—
1587	—	—	—	—	—	—	—
1603	—	—	—	—	—	—	—
1616	—	—	—	—	—	—	—
1620	—	—	—	—	—	—	—
1857	—	—	—	—	—	—	—
1863	—	—	X	—	—	—	—
1877	—	—	—	—	—	X	—
1917	—	—	—	X	—	—	—

Table 10. Pairs of papers and their coupling strength

635	pairs	coupled	with	1	coupling	unit
132	”	”	”	2	”	units
47	”	”	”	3	”	”
25	”	”	”	4	”	”
13	”	”	”	5	”	”
9	”	”	”	6	”	”
4	”	”	”	7	”	”
5	”	”	”	8	”	”
1	”	”	”	9	”	”
3	”	”	”	10	”	”
1	”	”	”	11	”	”
1	”	”	”	12	”	”

Those with n = 3 and 4 form another group of pairs. The third and fourth groups consist of pairs with n = 2 and n = 1 respectively. As a control we chose a random group of 100 pairs.* We then examined how each of the

976 pairs was regarded by the ASI method of indexing. We considered that if the pair shares one ASI category a relation exists. The results of this comparison are shown in Table 11.

Table 11. ASI relatedness of BC pairs.

Coupling strength of BC pairs	Percent of pairs related according to ASI
$n=5$ and over	86
$n=4$ and 3	71
$n=2$	67
$n=1$	55
100 random pairs ($n=0$)	0

The number of random pairs that can be formed from a population of N papers is $\frac{1}{2}(N^2 - N)$. In our case, $N = 334$ and the number of possible pairs is 55,611 Since the BC method produced 876 pairs, the probability of picking one such pair at random is less than two percent. We may, therefore, consider that for the one hundred random pairs n is virtually equal to zero.

Table 11 indicates that the coupling strength (n) is a measure of relatedness consistent with the judgment of ASI. The statement "Of the group of papers for which $n = 1$, 55 percent are related according to ASI" does not imply the corollary that the remaining 45 percent are not related. Of the remaining 45 percent, we can only say that they do not share an ASI category.

The comparison of Table 11 is asymmetric in its treatment of BC and ASI in the following sense. For BC we considered not only the fact of relatedness but also its relative strength. The judgment of ASI, on the other hand, was taken as a simple *yes* or *no*. Thus, we could count the number of *yes* and *no's* in each BC class of given coupling strength, but we could not in any given instance indicate a comparable ASI strength of coupling between any two papers. The results of the third stage comparison show that as the coupling strength (n) of the BC pairs increases, the probability of sharing an ASI category also increases.

• Fourth Stage

This experiment could not be performed because we found no consistent measure of strength of relatedness in the ASI method. Several attempts were made but all ended in failure. For example: Let a, b, c, d . . . be various ASI categories. Now consider two papers both categorized as $P(a)$, are they more or less related than two papers categorized $P(a,b)$ and $P(a,c)$? Or consider two papers, both categorized $P(a)$, are they more or less related than two papers, both categorized $P(a,b,c)$? The above and similar questions could not be answered in any consistent manner. This difficulty no doubt derives from the fact that the ASI categories are not of equal "logical size." We must bear in mind, however, that negative results are always tentative. We suspect that a close examination of the "logical size" of the ASI categories could yield a weighting factor that would give a measure of relatedness to the method.

• Summary

1. The papers in Vol. 112 of the *Physical Review* were processed by the method of Bibliographic Coupling and the results compared with those of Analytic Subject Indexing as performed by the editors of the *Physical Review*.

2. Groups formed by BC show high correlation with the verdict of the ASI method. Detailed results and numerical measures of this correlation are given.

3. Groups formed by ASI correlate with those formed by BC more or less, depending on the "logical size" of the ASI category.

4. The verdict of ASI confirms that the coupling strength of the BC method is a measure of strength of relatedness between pairs of papers. As the coupling strength goes up, the probability of relatedness by ASI also goes up.

5. No consistent measure of strength of relatedness between pairs of papers could be found applicable to the ASI method.

6. This report does not pass judgment on the utility of either method to any specific application.

References

1. M. M. Kessler, "An Experimental Study of Bibliographic Coupling Between Technical Papers," *IEEE Transaction PTGIT*, vol. IT–9, 1963, p. 49.
2. ——, "Bibliographic Coupling Between Scientific Papers," *American Documentation*, vol. 14, no. 1, 1963, p. 10.
3. ——, "Bibliographic Coupling Extended in Time: Ten Case Histories," *Inform. Stor. Retr.*, vol. 1, 1963, p. 169.

In the Department of Agriculture (DOA), the Director of the National Agricultural Library (NAL) has been assigned the responsibility for coordination of DOA science information activities.[6] A clearinghouse for bibliographies and translations was established at NAL.

Industrial Organizations and Professional Societies

The American Petroleum Institute (API) has developed a combined manual and computer-based information searching system for its abstracts based on three indexes: (1) a manual alphabetic subject index; (2) a manual "dual-dictionary" coordinate index; and (3) a supplementary search tape index. Experience has shown that 60 to 80 percent of all questions addressed to API are so specific that manual methods are more economical. The computer tapes are intended to serve the other 20 to 40 percent which include more complicated questions.

The CAS announced in November, 1964[12], that in January, 1965, it would inaugurate a new service, Chemical-Biological Activities Center (CBAC). This will combine a computer search facility with the biweekly index journal, *Chemical-Biological Activities*, which will cover 300 primary journals. Advance information will be provided CBAC by page proofs of the journals. This service will correlate chemical structure with biological activity. The search facility may be available late this year.

The American Society for Testing Materials has issued its CODEN[13] as a standard for abbreviation of periodical titles. The code for an abbreviation is a four letter combination such as JACS for *Journal of the American Chemical Society*. The National Academy of Sciences—National Research Council published the results of its survey on chemical notation systems[14] and a supplement to its report on Communication Problems in Biomedical Research.[15]

International[11]

The International Council of Scientific Unions Abstracting Board continued to assist its new members in the acquisition of proof copies and advance copies of primary scientific periodicals. The Board is considering the extension of its activities to include nuclear energy and geology.

The International Federation for Documentation (FID) has recommended that UNESCO assist in a survey of the *Bulletin Signaletique, Physikalische Berichte,* and *Referativny Zhurnal* similar to the one carried out on *Physics Abstracts*. It was also recommended: that (1) the International Standards Association (ISA) study the problem of abbreviations of titles of periodicals (such as CODEN), for use in input and output equipment; and (2) the scientific unions encourage the establishment of associations of editors similar to the Conference of Biological Editors.

The Training Committee, FID/TD, met in Warsaw in May, 1964, and recommended that chairs be established at universities for training documentalists and that there be an exchange program for documentalists. This Committee hopes to publish a world list of documentalist training schools and programs.

The Committee on Linguistic Problems, FID/LP, is engaged in a pilot project for the registration of new terms in the field of documentation. Editors of documentation journals have been asked to cooperate. FID is also preparing for UNESCO a multilingual vocabulary of documentation terms, and is cooperating with NSF in the preparation of an inventory of nonconventional automatic documentation systems in use.

UNESCO published the third edition of the *Handbook on International Exchange of Publications*. UNESCO also continued to encourage better coordination between periodicals and abstracting and indexing publications and to encourage organizations and institutions to execute agreements on standardization in abbreviations, transcriptions, references, etc. In connection with the latter, UNESCO gave a contract to the FID for the preparation of a list of documentation terms in English, French, German, and Russian. UNESCO will continue to promote the establishment and strengthening of national or regional scientific and technical documentation centers through training courses and seminars.

UNESCO Working Party No. 3, on scientific translation and terminology, met in Rome in January and February, 1964. It recommended improved coordination of terminology, lexicographical activities and the publicizing of ISA's work in this field; a study on the need for additional translation centers; and collaboration between UNESCO and other organizations for the production of a directory of scientific translation services.

The Office for Economic Cooperation and Development (OECD) continued publication of a series of reports on research facilities of the member countries. Bibliographies were published for various industries such as glass and ceramics, fuel and heat processing, pharmaceuticals, etc.

The Academy of the Rumanian Peoples Republic established The Scientific Documentation Center in January, 1964. The Center will provide Rumanian scientists with domestic and foreign scientific information, including bibliographies, and will do research and development in documentation and information systems. The Center published *Rumanian Scientific Abstracts* in two series, Social Sciences and Natural Sciences. The *Abstracts* are published monthly in English and in Russian.

The Zentralinstitut für Dokumentation in East Germany was established under the State Planning Committee on October 1, 1963 (successor to Institut für Dokumentation der Deutschen Akademie der Wissenschaften). It is responsible for coordinating and organizing the various services into a uniform and effective documentation system.

The Soviet Union and other countries of Eastern Europe held a conference in February, 1964, to discuss standards for library construction and equipment. It was decided to establish a center for information of this type at the State Library of Czechoslovakia.

● Research and Development in Scientific Communication and Documentation

The COI has depended regularly upon the publication *Current Research and Development in Scientific Documentation* (CRDSD) of the NSF for data for its report on R&D in the information field. CRDSD is prepared by the OSIS of NSF from solicited information reports and usually appears on an annual basis. The latest, No. 13, covers data collected during the summer and fall of 1963.[16]

In compiling CRDSD No. 13, NSF collected 495 statements from 380 organizations, of which 166 were statements from 126 organizations in 27 foreign countries. By count from the 495 statements, there were 599 projects identified by COI.

Analysis of the data provided by CRDSD No. 13 and the previous reports of COI[1,2] is shown in Fig. 1. The data are plotted as near the end of the data collection period in each year as could be determined for each CRDSD. In previous reports the data were plotted as of the date of publication of each CRDSD. The new curve is thought by COI to be more representative of the actual state of affairs.

Figure 1 shows the number of projects reported in various CRDSD's from 1957 to Fiscal Year 1963 for the world, broken down into the number of United States and foreign projects. The total number of U. S. projects is further broken down into those paid for by U. S. Government money and those paid for by private money. Another curve shows the number of U. S. projects reported which were executed by private organizations (whether paid for by Government or private money).

Figure 1 indicates that the curve for R&D in documentation was still rising in mid-1963 but not as steeply as before.* The curve for foreign projects showed an increase in steepness but this probably reflects better reporting rather than an increase in the number of projects. All of the other curves showed a lower rate of increase than did the world curve.

In spite of their qualitative nature, the CRDSD data offer some interesting information about: (1) the kind of R&D done in the U. S. and abroad; (2) the emphasis put on the different kinds of projects in the U. S. and abroad; (3) the approximate degree of change in the types of research which has taken place since the last report; (4) the type of organizations doing the research; (5) the government organizations which sponsor research; and (6) the research done in foreign countries.

Types of R&D in the U. S. and Abroad.

A comparison of U. S. and foreign projects is given in Table 1, using the CRDSD categories. In the U. S. the greatest emphasis was on information storage and retrieval and on "related research." Each accounted for 36 percent of the total number of projects. Greatest foreign interest was in related research (41 percent of projects), information storage and retrieval (28 percent), and mechanical translation (21 percent).

The U. S. indicated greater interest in research in information needs and uses and in artificial intelligence than did foreign countries. The foreign projects, on the other hand, reveal greater interest in mechanical translation and in linguistic and lexicographic research than was true of the U. S.

The computed "change factor" given in the table indicates that in the U. S. about 40 percent of the reported

* The curves of Fig. 1 do not precisely duplicate those of previous COI reports; however, the overall curve entitled "Total Projects: World," when compared with the overall curve of Fig. 1 in the 1962–63 report, does appear to show the rounding off or beginning of an apex, as forecast in Part V of the 1962–63 COI report.

Table 1. Kinds of research done in the United States and foreign countries and degree of change between CRDSD Nos. 11 and 13.

	Category	U. S. projects %	U. S. projects Change factor*	Foreign projects %	Foreign projects Change factor*
1.	Information Needs and Uses	19	1.6	8	0.5
2.	Information Storage and Retrieval	36	1.3	28	1.8
3.	Mechanical Translation	5	0.6	21	1.5
4.	Equipment	5	1.0	2	0.5
5.	Related Research	36	1.2	41	1.5
5.1.	Character and Pattern Recognition	(9)**	1.0	(9)	1.9
5.2.	Speech Analysis and Synthesis	(5)	1.1	(5)	1.3
5.3.	Linguistic and Lexicographic Research	(7)	0.9	(19)	1.5
5.4.	Artificial Intelligence	(12)	1.8	(7)	1.4
5.5.	Psychological Studies	(3)	1.3	(1)	1.0

* Change factor indicates greater or less activity than shown by CRDSD No. 11. A factor of 1.0 indicates no change. CRDSD No. 11 covered the period immediately before that covered by CRDSD No. 13.
** Numerals in parentheses are subtotals for Category 5.

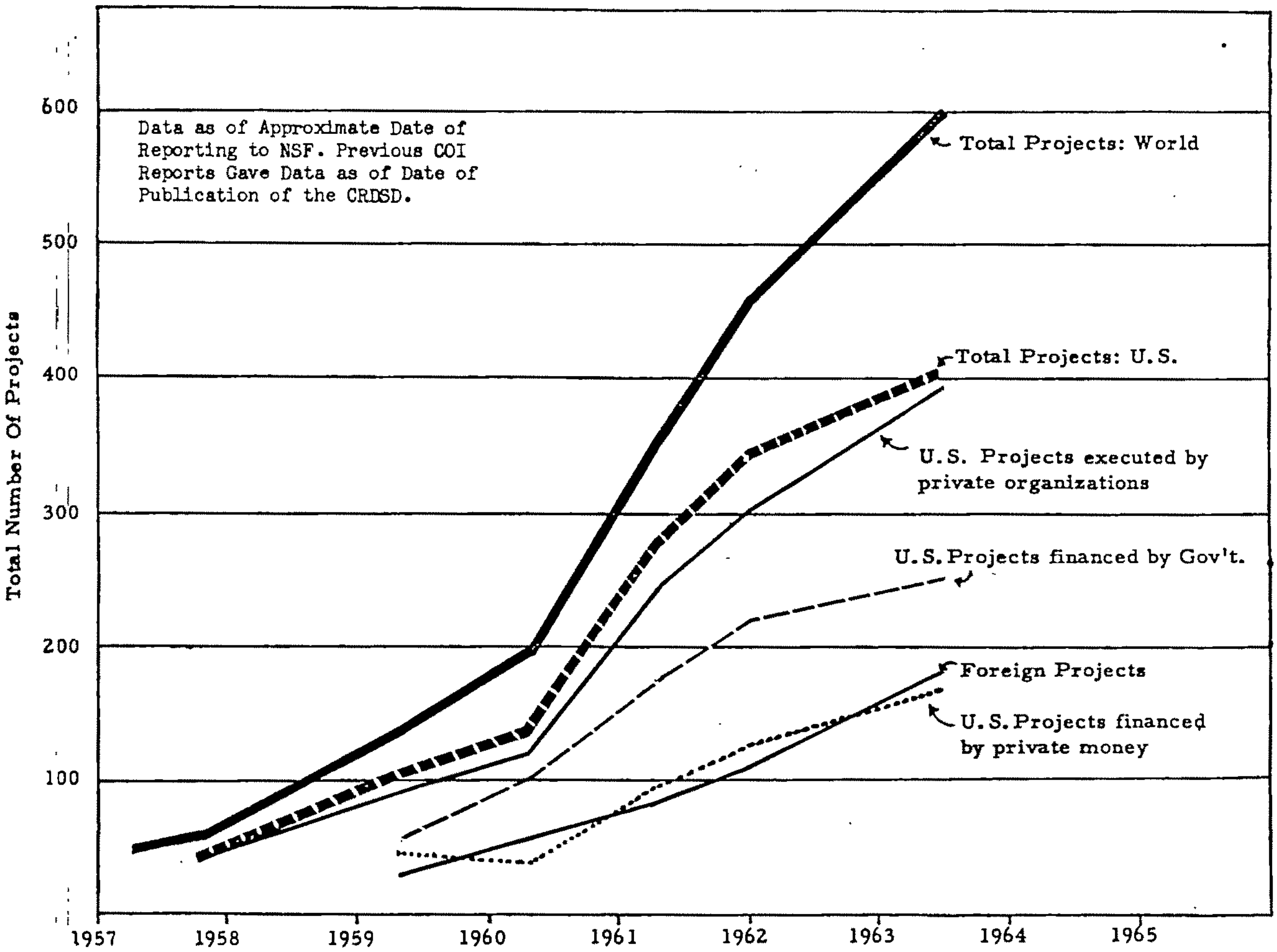


FIG. 1.

machine translation projects were ended or dropped during the period. About 50 percent of the reported foreign projects in information needs and uses and in equipment ceased.

In the U. S. there were about 60 percent more projects in information needs and uses, 30 percent more in information retrieval, and 80 percent more in artificial intelligence than in the previous period. In the other categories there was little change, or the percentage change was insignificant because of the small number of projects. Among foreign projects, gains of 30 to 90 percent were made in all categories except two which decreased and one which remained unchanged.

Types of Organization Doing Research

Analysis of the 495 statements to determine the kinds of organizations doing the research produced the results shown in Table 2. Industry and the universities shared about equally in total programs, together accounting for over 70 percent. Of the others, 10.8 percent was carried out by research institutes and foundations, and 8 percent by government research groups. Societies and information groups (libraries, documentation centers, etc.) each had less than 5 percent.

There seemed to be no anomalies in Categories 1, 2, 3, and 4. Half or more of the projects were carried out in industry or the universities, and the distribution among the others does not appear unusual. As was expected, over half of the machine translation work was done by universities and the bulk of the equipment fell to industry.

It was startling, however, to find that industry was doing as much of the research in Category 5 as were the universities, since it might be expected that the universities would dominate this area. Between them they accounted for nearly 80 percent of the research in area 5.

A comparison is given in Table 3 of the three types of organizations which carried out most of the programs in area 5. Industry carried out 63 percent of the programs in character and pattern recognition (5.1) and nearly half of those in speech analysis and synthesis (5.2). This seems an unusually high proportion for industry. The universities had nearly half the programs in linguistics and lexicographic research (5.3), and research institutes (linguistic mostly) followed with about

Table 2. Types of organizations doing research. (Taken from CRDSD No. 13)

Organization	Category 1	2	3	4	5	5.1	5.2	5.3	5.4	5.5	Total	%
Government Research Groups	7	17	5	3	8	5	1	–	2	–	40	8.0
Industry	15	61	7	18	78	34	14	8	21	1	179	35.9
Information Groups	6	14	1	–	2	1	–	1	–	–	23	4.6
Research Institutes and Foundations	4	13	8	2	27	5	2	16	3	1	54	10.8
Societies	7	5	–	–	2	–	–	2	–	–	14	2.8
Universities	15	42	29	1	85	8	14	28	24	11	172	34.5
Others	4	7	1	–	5	1	–	3	1	–	17	3.4
Total Statements	58	159	51	24	207	54	31	58	51	13	499	
Total from CRDSD No. 13											495*	

* The difference is accounted for by joint contracts or subcontracts on some programs.

28 percent. Industry and the universities shared almost equally in the bulk of the programs concerned with artificial intelligence (5.4). The universities conducted most of the psychological studies (5.5).

Location of Foreign Projects

To the COI's knowledge there has been no published tabulation of the geographical location of foreign projects. Since it is believed that such a tabulation is useful, it is supplied in Table 4.

Most of the 182 projects reported were in Western Europe (115); Eastern Europe reported 40. It should not be concluded that Czechoslovakia did more R&D than did the U.S.S.R. because it reported more projects. The projects may have been smaller. In the Far East, Japan accounted for 18; and India, Argentina, Canada, Colombia, Israel, and Mexico for the other 9.

Table 3. Who does research in potentially related fields by percentage. (From Data in CRDSD No. 13)

Organization	Category 5.1	5.2	5.3	5.4	5.5
Industry	63	45	14	41	8
Research Institutes	9	7	28*	6	8
Universities	15	45	48	47	84
% of Total Programs	87	97	90	94	100

* Many linguistic research institutes and groups.

• U.S. Government Sponsorship of Research

Among the 599 projects reported by CRDSD No. 13, 250 were paid for by the U. S. Government. Breakdown by agency is shown in Table 5. The sponsors reporting the greatest number of projects were DOD, NSF, and HEW in that order.

Table 4. Geographical distribution of 182 foreign projects. (From CRDSD No. 13)

Geographical area	Countries	Number	Totals
Western Europe	Great Britain	40	
	West Germany	24	
	France	18	
	Italy	11	
	Sweden	7	
	Netherlands	7	
	Belgium, Finland, Austria, Yugoslavia	8	115
Eastern Europe	Czechoslovakia	16	
	USSR	11	
	Hungary, Rumania, Poland	13	40
Far East	Japan	18	
	India	2	20
All Others			7

Table 5. Government sponsorship of research and development projects reported in CRDSD No. 13.

Agency	No. of projects	Per cent
Agriculture	3	1.2
AEC	4	1.6
Commerce	6	2.4
DOD	122	48.8
HEW	40	16.0
Interior	1	0.4
NASA	4	1.6
NSF	69	27.6
Post Office	1	0.4
Total	250	100.0

It proved impractical, because of lack of a common basis, for COI to compare the CRDSD data with NSF's data for the fiscal years 1962, 1963, and 1964.[17] In Part II of the NSF publication is found a category 4 which

is defined as "the conduct of research and development on techniques and machines for improving scientific and technical information tasks. This category also includes the conduct and support of studies and surveys to identify scientific information problems."

The data reported under category 4 were used by the Bureau of the Budget and is cited by the Elliott Committee.[7] These data are used as another measure of the Government sponsorship in this field.

Trends

1. There are signs of a leveling off in the rate of increase of information research activity. It is subject to debate whether this is to be taken as a sign of the coming of age (end of Phase II) of "information science" as suggested in our 1962-63 report.

2. The message of the Weinberg report that information transfer is inseparable from the research process is apparently being taken more seriously. The Director of Defense Research and Engineering named 22 information analysis centers as proper line items of the R&D budget. He carefully differentiated such centers from libraries and documentation centers whose function is primarily document retrieval. It is reported that eleven more analysis centers are being considered.

3. The trend away from large monolithic centers and toward those with more flexible functions continues. Examples are the reorganization of the Defense Documentation Center (DDC) and the establishment of CFSTI. CFSTI will serve as the outlet for unclassified reports with unlimited distribution.

4. There is a continuing trend in the desire for compatibility of such items as equipment, systems, codes, and thesauri. There is also a countertrend. Compatibility is opposed by some of those who are continually developing new equipment and systems. Standardization will probably come more easily in microimage and copying equipment, where some standards already exist.

5. The trend continues in cooperation, simplification, and coordination of Government information activities with establishment of focal points both within and without departments and agencies. Examples are the previously referred to establishment of CFSTI in the Department of Commerce and of information and data analysis centers in the DOD.

6. Both discipline-oriented and mission-oriented services and information centers continue to develop. The Chemical Abstracts Service is an example of the first. The subject matter with which it deals is well adapted to data processing. Examples of the mission-oriented services are many of the information and data analysis centers of DOD, which are problem-oriented.

7. If the copyright bill introduced in the House and the Senate in July, 1964, becomes a law it may influence trends in duplication and copying as a means of communication.

8. Education for information science has so far made but small strides. The reasons undoubtedly lie in lack of agreement on basic needs for undergraduate and graduate programs. A great deal of experimentation will be needed, and theory must be developed.

References

1. Committee on Organization of Information, American Documentation Institute, "Report for 1961–62," *American Documentation*, vol. 14, no. 4, October, 1963.
2. ——, "Report for 1962–63," *American Documentation*, vol. 15, no. 4, October, 1964.
3. ——. Replies to requests for information from about 30 leading universities; taken from the COI files.
4. S. M. Isabella, "Education for Information Center Personnel in ALA Accredited Library Schools of the U. S. and Canada," master's thesis, Drexel Institute of Technology, Philadelphia, Pa., 1964. See also *Proceedings of the American Documentation Institute*, vol. 1, Spartan Books, Inc., Washington, D. C., October, 1964, pp. 67–71.
5. R. S. Taylor, "Curriculum Development in Documentation and the Information Sciences," *Proceedings of the American Documentation Institute*, vol. 1, Spartan Books, Inc., Washington, D. C., October, 1964, pp. 31–37.
6. Federal Council for Science and Technology, Committee on Scientific and Technical Information (COSATI), "Summary Progress Report," Washington, D. C., September, 1964.
7. U. S. Congress, "Report of the Select Committee on Government Research of the House of Representatives," Carl Elliott, Chairman. 88th Congress, 2nd Session, Washington, D. C., November 20, 1964.
8. Department of Defense Instruction No. 5100.45, July 28, 1964.
9. G. S. Simpson, Jr., "Discussion of Functions of Information/Data Analysis Centers," The Institute of Data/Information Availability, The American University, Washington, D. C. (unpublished).
10. Gilbert W. King, et al., "Automation and the Library of Congress," Library of Congress, 1963, 88 pp.
11. OSIS of NSF, *Science Information Notes*, 1964 (various issues).
12. "CAS Launches CBAC," CENE, vol. 42, no. 46, November, 1964, p. 64.
13. Lester E. Kuentzel, Ed., "CODEN for Periodical Titles; An Aid to the Storage and Retrieval of Information and to Communication Involving Journal References," prepared and maintained for Committee E–13 by the Wyandotte-ASTM Punched Card Project, Philadelphia, Pa., 1963, 1st Supplement, 1964.
14. "Survey of Chemical Notation Systems," NAS–NRC Publication 1150, Washington, D. C., 1964.
15. "Communication Problems in Biomedical Research," supplement to a report submitted October 31, 1963, to the Director, NIH, NAS–NRC, March 10, 1964.
16. NSF, "Current Research and Development in Scientific Documentation (CRDSD), No. 13," November, 1964.
17. Federal Funds for Research, Development, and other Scientific Activities. NSF, November, 1964.

BOOK REVIEWS

7/65-1R **SYNTOL: Volume II.** 1965. J. C. Gardin. Graduate School of Library Service, Rutgers, the State University (Rutgers Series on Systems for the Intellectual Organization of Information, Susan Artandi Ed.), New Brunswick, New Jersey.

The seminar papers being published in this series had a prescribed order beginning with general information and including the system's name, presumably with its derivation, and its type. But SYNTOL's obviously synthetic name is not explained until p. 43, where it is spelled out as Syntagmatic Organization Language. It is not typed as a retrieval system until well over halfway through, where it is said that, "as one might expect, the retrieval method with SYNTOL is basically that of co-ordination" (p. 67). A note to this says, however, that "to the extent that pairs of related descriptors are explicitly recorded in the file, the system can be conceived as pre-co-ordinative" (p. 76). As one might expect from this, its syntagmas are "pairs of related descriptors."

Some form of coordinate indexing might have been expected from earlier statements that "it is a computer-oriented system [but] the devices required . . . vary according to the level of sophistication which one may choose to adopt in its application. The minimum mechanism is a 'collator' of any kind, peek-a-boo' card index for example." (pp. 13, 17, 75). But though coordinative, it is not a uniterm system, since its units are the binary or dyadic syntagmas. Other terms from other systems are freely used, such as *descriptor, facet, thesaurus,* but *syntagma* seems its own and its novelty—though with some doubts arising, apparently after discussion at the Rutgers seminar from which the paper as published is the outcome. It is in this way that the seminars and the series are serving a very useful purpose, not of guaranteeing systems but of shedding light on them, even though Rutgers and the editor, Dr. Susan Artandi, can hardly be expected to dispel all obscurity.

There is an admission that the suffix L in SYNTOL "perpetuates an overuse of the word 'language' in the field of information storage and retrieval" (p. 43). But the prefix SYNT, which is from the linguistic term syntagma, might be more questionable than the general term language.

In the linguistic theory and terminology to which SYNTOL is related, a syntagma is a "meaningful unit of word association," but not necessarily SYNTOL's two-term unit; for example, Drugs . . . Animals, in some implicit or explicit relation. And in its indexing use, *syntagma* seems to be no more than one more esoteric synonym (pp. 28, 45). Indexing has to have its own grammar or tyntax, different from that of the sentence which has neither the purposes nor the limitations of indexing, and an attempt at a relation of one to the other seems to introduce more confusion than clarity. The indexing idea of SYNTOL's two-term indicators seems to be that with unit terms there is more often than not a roles and links problem, which can be anticipated by precoordination. But the role problem remains, insofar as a mere relation or juxtaposition of two terms does not clear up ambiguities which there may be in the relation.

SYNTOL is French, but seems to have some derivation from the British facet analysis school, and especially on relations from Vickery and Farradane. Four "syntactical relations" are distinguished, called *predicative, associative, consecutive, coordinative.* Here again emphasis is on language instead of logic, and as with the facet classification there is little or no consideration of the fundamental logical relation of object and aspect, which Kaiser specialized as Concrete and Process. Finally after much algebraic formulation of R. 1-4, this categorization of relations is not insisted on, more reliance being placed on "the syntagmatic context (i.e., the two related descriptors) acting as an effective discriminator" (pp. 37–38, 40, 47, 54–55).

What might be called the object of the exercise seems to have been a basic logic or grammar of indexing in natural language, restricted as it must be in indexing use, with "a symbolic language for the manipulation of expressions belonging to the . . . natural language," with the incidental apparatus of a subject headings or descriptors list, called a lexicon, and a thesaurus of natural expressions with descriptor equivalents "where the correspondence is not self evident" (pp. 11, 52–53).

The system, like others, seems to have been developed with some, but not much or prolonged, application, especially at the computer level. It is set out in its own terms or jargon, with a graphical algebra which seems to have no more relation to indexing essentials or to coding for mechanized retrieval than its supposed linguistic base. Then, after presentation at a Rutgers seminar, its author, who is an archaeologist, seems to have tried not only to improve its representation in English but also to relate it to American thinking and terminology (somewhat divergent from European) and especially the thinking and terminology of coordinate indexing, with its later roles and links. In this, he is deserving of some sympathy. He seems also to have been obliged to argue for indexing, as opposed to machine searching of texts directly, having apparently begun by assuming human indexing with mechanization of storage and retrieval without much thought of direct machine retrieval from natural language texts (pp. 12, 87). But on this he is entitled to say *show.*

There is, as he says, a multiplicity of systems and a need for a fundamental approach, but perhaps there is more a multiplication of theories and terminology than of essentially different systems, and the remedy at any level of approach may not be yet another questionable excursion into questionable theory, with yet another terminology. There is a glossary of six special terms but with reference to another one of 160 (p. 102). This is only half of the 320 of Ranganathan's classification terminology (his Prolegomena, 1957, p. 167), but enough.

John Metcalfe
Director, School of Librarianship
The University of New South Wales

7/65–2R **Evaluation of the Impact of a Citation Index in Physics.** 1964. Ben-Ami Lipetz. American Institute of Physics, New York. AIP/DRP CI–3–1964.

Citations of Physics Literature. (Experimental issue) 1964. Ben-Ami Lipetz. American Institute of Physics, New York. AIP/DRP– CI 1964.

Every new bibliographic tool probably generates a certain amount of controversy when it is introduced and in its early

stages. The citation index is no exception. As a matter of fact, it has probably caused more than the "normal" amount of heated discussion. Thus, it is encouraging to read a report like this which aims at an unbiased, quantitative evaluation of a citation index. It is an area where objectivity is badly needed.

The two volumes of the Lipetz report resulted from a survey he made in which the object was to measure the value to a select group, probably mostly physicists, of a special citation index which was given to them at no cost and with no fanfare. In the index were all the articles in four translated Soviet physics journals which cited references from two leading American physics journals. The plan was to measure the use made of the four Soviet translated journals before and after the citation index was sent to the test group, which was kept unaware of the survey going on. The survey measured browsing activity, borrowing, photocopying, placing of subscriptions and ordering of back volumes of the four test journals. Measurements were made in the geographical test area at the same time that they were made in other locations in order to make a comparison. About 60 libraries outside the test area and 25 libraries in the test area took part. Indexes were sent to 553 subscribers of one or both of the American physics journals who happened to live in the test area. Questionnaires and some personal interviews were used to learn their reaction to the index supplied each of them. About 260 replies from recipients were received.

The results of the library statistics showed a definite but small increase in the use of the four Soviet journals in the test area compared to the use in the area where indexes were not sent. Those who used the index were mostly favorable to it, but most of the recipients did not actually use it. Of the 260 replies, there were 122 favorable responses, but it should be noted only 38 of those favorable to it had used the index, the rest merely looking at it.

The results seem likely to be interpreted in two ways. The proponents of citation indexes would be pleased at the fact that the test area saw an increase in usage and that those who used the index favored it by a large majority (38–4). On the other hand, those doubtful about such indexes would point out that the author describes the increase in use as "quite small," and that only 38 of the favorable replies came from actual users, thus placing the balance of the 122 favorable replies in a rather uncertain position, because they were from persons who had only looked at the index, not used it.

At any rate, the impartiality of the survey by Mr. Lipetz is quite welcome as a serious effort to evaluate citation indexes. One fact that may have harmed the response was to format of the index used. Its cited articles were arranged chronologically by the order in which they appeared in the American journals, rather than by an alphabetical order of cited authors. This was done to save money, but was it a fair presentation of a citation index?

It is hoped that this is only the beginning of research in this area. One aspect that many would like to see tested would be the relative acceptance by both scientists and librarians (and related groups) of citation indexes versus more traditional indexes, such as author-subject bibliographies, KWIC permuted title indexes, etc., if a way to cover the same literature, such as the four Soviet journals, could be found to make the comparison more meaningful. The Lipetz report shows acceptance for a citation index when none other is available. What would happen when the test group has several indexes to evaluate, all on comparable literature, recognizing the basic differences between citation indexes and those for retrospective searches?

Ellis Mount
Engineering Librarian
Columbia University

7/65–3R **Catalog Card Reproduction; Report on a Study.** 1965. George Fry and Associates. American Library Association (Library Technology Project Publication #9), Chicago. 81 pp.

It is part of the "black box syndrome" that computer systems and automatic output may seem to obviate catalog cards in the very near future. Nevertheless, cards have inherent advantages which will assure them a substantial place in libraries and documentation centers for the foreseeable future.

This small volume is designed to provide the basis for selection of a process for reproducing catalog cards. Unfortunately, it does a disappointing job. The report is of some use if one wishes to have cost methodology suggested. It provides illustrations of 20 pieces of equipment which can duplication cards but does not say which model. It includes 16 interesting illustrations of the quality of card reproduction. Yet it seems typical of this volume that these quality illustrations are based on different originals, different typewriters and different type faces so that true comparison is not possible; neither of the two enlargements is of one of the cards illustrated.

The manual is divided into sections which consist of general description and discussion of equipment and procedures, comparative cost tables, a method by which cost comparisons between various processes can be obtained, and general equipment recommendations. Processes covered are: purchase of cards, standard electric and automatic typewriters. fluid duplication, postcard size and full size stencil duplication, offset with a variety of methods for preparing masters, fiber stencil and metal plate addressing machines, diazo duplication, electrostatic, diffusion transfer, and projection photocopying.

Although subtleties are lost in abstracting recommendations, it is clearly stated that fluid or small stencil duplicators are best when volume is below 2,000 titles per year. Above that level, full size stencil duplication will be more economical. At a level of about 9,000 titles, the advantage changes to offset equipment with electrostatic preparation of the master. It is noted, however, that, even in large volume, electric typing is most efficient for titles having brief description and requiring only three or four cards. These recommendations parallel the report by Joseph H. Treyz, "Equipment and Methods in Catalog Card Reproduction," published in *Library Resources and Technical Services* (vol. 8, p. 267).

In judging this volume, it is difficult to determine the audience to which it was aimed. The report does not have sufficient detail to really assist a major library in discriminating among pieces of equipment or in refining its operation. It seems more probably that it is aimed at a novice in a small library who is faced with the task of first duplicating cards by other than typewriter. Even here one looks in vain for mention of the myriad practical details that will make any rude process an effective system; e.g., the simplest stencil system can use a number of different stencils and different fast-drying inks, and on this point the report makes the sage recommendation that the operator "should try several of these . . . until he finds the best one for his operation."

Concerning the fine points of card duplication, there seems to have been no attempt to appraise the comparative advantages of, for instance, the A. B. Dick Model 350 with the Addressograph-Multigraph Model 1250. There is no attention to the quantity of waste cards due to improper inking or adjustment of image on first cards of each set. There seems to have been no test of the reliability of card feeding or consideration of which machines can provide for dual track feeding. No discussion is provided of which machines save steps and arm motion and thus lessen operator fatigue or which have more convenient controls for adjustments. There is no attention to which are easier to convert if sheets (for library bulletins, acquisition lists, etc.) are to be produced on the same equipment with catalog cards. No discussion is provided of which are more economical, convenient, and clean in their use of supplies. No comparison is given of the normal life of machines used for card duplication. There is but slight consideration of the service experience with different machines, although in a small library as much as 5 percent of the total cost of card reproduction may be in service and parts.

With the $49,470 that went into this study, one wonders if it might have been advantageous to use a case study method for better presentation of detail and comparison. Indeed, since several articles exist on individual applications, it seems curious that no reference is made to those which make some contribution to this subject. Although this Chicago management consulting firm visited some 70 libraries to investigate local card duplication processes, there is not list in the book to lead one to pay special attention to the application of the Ektalith process which has had successful use at the San Francisco Public Library, or to the Itek Platemaster used effectively at the University of Texas Library, or the Copyflo process used for over two years at Harvard University Library.

Finally, one is struck by the fact that there is no consideration of the effect of the number of added entries per set, that is, the number of unique headings that must be added at the top of the card. (The only references in the report to this factor are mention of Wilson's preprinted sets and comment that the stencil overlay technique permits a heading change.) Yet there is a marked economic difference when a title receives one or two main entries with short descriptive information and a dozen added entries, as may be the case for a technical report, as compared to the situation when a dozen uniform author cards are suitable for works with non-distinctive titles where copies are placed in a dozen laboratory or branch libraries. For how many copies of one added entry is an overlay justified? In what circumstances is the Flexowriter advantageous?

Let no one underestimate, however, the difficulty of appraising the many variant processes of card duplication. The study was necessarily a large one. The report is well written, well indexed, and handsomely printed. It pulls a good deal of information together for convenient use as a compendium; and it can assist one to eliminate consideration of those processes which are inappropriate to the local situation. If the reader is attentive, the report reveals significant variables and points out defects to be guarded against, such as the impermanence of impression in a few processes.

Nevertheless, it is sad that as far as providing definitive advice for the setting up of a system, improvement of procedures, and the purchasing of particular equipment, the manual fails. No library can turn to this report for exact guidance in selecting a piece of equipment which will best do its job and in working out the particulars of its system.

David C. Weber
Stanford University Libraries

(b) Even if these objections are rejected, there is still this: What is needed (if we aim at true hospitality in array) is not simply a means of *extending* the coordinate series, but of *intercalating* coordinates between extant coordinate classes.

If *.1—.8* is systematically arranged, and need arises to intercalate between *.1* and *.2*, it does us no good to be able to add *.91* as coordinate, because this divides the array at the wrong point. Nor can we use *.15*, since this a member to the array subordinate, not coordinate, to *.1*. However, the use of a different punctuation (I would suggest the comma) would open the way to easy systematic intercalation of any desired generic/specific level, in a way far easier to comprehend (in an optically scanned system) and unnecessary to program around (in an electronically scanned system).

If our original scheme were arranged thus:

new coordinate classes at any level could be intercalated by use of commas:

which could of course themselves be subdivided further:

Between the original intercalations and the original order new intercalations could of course have plentiful space:

until need arose for repetition of the comma as inter-intercalator:

By use of such a device the proposed U D C change of main class *4* to indicate classes *5* + *6* would be rendered unnecessary; *5,5* would be superior in several obvious ways. By use of the comma device, the L C coordinating intercalator (the decimal-point) could become the comma, and decimals could become *sub*ordinating devices there as well.

JEAN M. PERREAULT
Head, Information Retrieval Division
Florida Atlantic University
Boca Raton

SPECIFICATIONS FOR A SUPERIOR MICROTEXT READING MACHINE

Until this decade, users of microtexts worked with good off-the-shelf machines which cost from 300 to 900 dollars. In the past five years, the engineering drawing field has substantially advanced the capabilities of reading machines. Machines suitable for research libraries have, unfortunately, made almost no progress, and the photographic industry seems to have ignored libraries.

The Council on Library Resources which was set up in 1956 to strengthen American libraries—especially research libraries—has apparently been working for the lone scholar located in a relative wilderness of research instead of trying to meet the substantial needs of the major centers of research. The CLR has for seven years supported the development of a small, inexpensive and portable microfilm reader.

What research libraries need is not a better Volkswagon, not more Chevrolets or even Pontiacs, but a few Cadillacs for scholars and students to use in conducting their research. Such a Cadillac can be built today and its requirements are here described.

It has been determined many times that the users of libraries are dismayed by the need to use microfilm. Even today most scholars are instantly repelled by the thought of having to use film. When offering a microfilm in lieu of the original, many a librarian has found his faculty unwilling to forego buying a journal or against disposing of the space-consuming and seldom-used opera of a minor pedant. There is a strong, conscious opposition to the microform.

One very sound reason for such opposition is the lack of any really good, efficient, comfortable, convenient, versatile reading machine. While some individuals would be pleased to use a small portable reader on an airplane or at a lounge chair in their studies, every researcher, scholar and student would be impressed with an pleased to use a large non-portable reader having all the convenient and efficient features of an H. G. Wells or Buck Rogers fable of the future.

Machines now on the market fall far short of being satisfactory. Not one presents a well designed environment for the lengthy and meticulous scholarly use common in research libraries. The situation resembles an automobile designed without regard to the driver's manual ease, his versatility of control, the visual conditions, and functional comfort. Beyond that, even the better machines, such as the 3M Filmac 200–R or the Itek 18:24 Reader Printer, fail in such fundamentals as rotatability of the projection head, variable enlargement, and electric drive.

There are several conditions which could make a Cadillac style of microtext reading machine practical today. First there is the technical feasibility. Fortunately the present state-of-the-art of optics, viewing screens, and mechanical controls makes the proposed machine feasible at once.

A second condition might be the limited budgetary capacity for equipment in research libraries, for such superior machines will cost much more than libraries are used to paying. This is less of a concern in 1965 than in any past year, since more special funds for one-shot purchases now exist than ever before, e.g., foundation grants, NSF funds, Woodrow Wilson grants, NDEA funds, the ACRL program, and so on. Budget balances at the end of a fiscal year are good sources; faculty members have research project funds that are available; and more federal money and industrial contributions are available to schools and colleges through greatly expanded programs in aid of education.

A third condition might be the dubious national market, since the one hundred major research libraries would probably each purchase only a few such machines in this decade. Actually the 2,000 libraries of academic institutions would purchase a significant number. Furthermore, the 16,000 governmental libraries, public libraries, and the special libraries of industry and commercial organizations would be a very much larger market. Taken together these institutions certainly constitute a market of sufficient proportions to be served.

When a *markedly* superior machine is available, libraries will rush to make it available to their readers, as evidenced by the phenomenal acceptance of the Xerox 914 copier. Scientists and engineers who need highly specialized laboratory equipment that may be used for a brief period are well supported in this society. Similar support is beginning to come for humanistic scholars and social scientists for whom such superior microtext equipment would be a major aid constantly used.

The proposed machine must meet the ASA specifications and treat satisfactorily the points of Ballou and Hawken. It should offer the best features of modern radar screens, of language laboratory controls, and of electronic computer packaging. It should be designed as an exemplar. A superior new reading machine for use in information centers, libraries, and similar organizations would take microtexts into a new generation of acceptance and use.

The intent of the general specifications which follow is to present environmental conditions vastly better than exist today and technical capabilities substantially better than are now combined in any reader:

1. The machine should be a floor model built into a study table with an overall 36″ x 60″ writing surface. The table should have acoustically treated sides and back extending about 20″ above the writing surface with an 8″ wide shelf on the back 14″ above the writing surface. An upholstered swivel chair should be part of the equipment.
2. All control mechanisms (focus, zoom, advance and reverse) should be in the lower front of the equipment. The viewer should be engineered so that it is efficient to use, and obvious in its method of use, in order that it be encouraging to the reader. This requires use of instructional labels, succinctly worded and legibly imprinted.
3. It should accommodate self-threading 16mm. and 35mm. nonperforate 100 foot rolls of film. It should be able to handle microfiche up to 9″ x 12″ on a stage with a frame-locating device consisting of a single knob or lever. (It would be desirable for opaque sheets to be accommodated, but it is not necessary if

NFSAIS Plans for Coordination

Representatives from major U. S. information services gathered from March 24 to 26 in Columbus, Ohio, to analyze the interfaces between discipline- and mission-oriented and between government and management approaches to abstracting and indexing services. Major emphasis was placed on the fields of biology and chemistry.

During panel discussion moderated by Verner Clapp, President of the Council on Library Resources, a plan was presented by Bernard Fry, Director of the Clearinghouse for Federal Scientific and Technical Information of the National Bureau of Standards, for the creation of four clearinghouses for the wholesaling of abstracts from profession-oriented abstracting and indexing services to project-oriented services or directly to users. These clearinghouses would be divided into the following areas of science: (1) agricultural, (2) engineering, (3) physical and chemical, and (4) biomedical.

Other participants at this meeting included: Scott Adams, Deputy Director, National Library of Medicine; Larry X. Besant, The Chemical Abstracts Service; Joseph F. Caponio, Scientific and Technical Communications Officer, National Institutes of Health; George A. Doumani, Head, Cold Regions Bibliography Section, Science and Technology Division, The Library of Congress; Paul Feinstein, Assistant Director, Scientific and Technical Information Division, National Aeronautics and Space Administration; William C. Hoida, Research Coordinator, Biological Abstracts; Keatha K. Krueger, Scientific Communications Officer, National Institute of Health; Marjorie R. Hyslop, Associate Director, ASM Documentation Service, American Society for Metals; Ljubo Lulich, Chief, Pesticides Information Center, U. S. Department of Agriculture; Phyllis V. Parkins, Director, Biological Abstracts; John Sherrod, Chief, Information Services and Systems Branch, U. S. Atomic Energy Commission; and Isaac D. Welt, Director, Cardiovascular Literature Project, The American University.

Plans were further advanced for restructuring the Federation into a more dynamic, forceful, and broadly based organization. The following persons were elected as officers:

President: Carolyn M. Flanagan, General Manager, Engineering Index, Inc., New York, New York
Vice-President: Foster D. Smith, Jr., Director, Science Information, American Geological Institute, Washington, D. C.
Secretary: Ralph E. O'Dette, Director, External Affairs, The Chemical Abstracts Service, Columbus, Ohio
Treasurer: Phyllis V. Parkins, Director, Biological Abstracts, Philadelphia, Pennsylvania

Abstracting and Indexing Service
RAYMOND A. JENSEN
National Federation of Science
Washington, D. C.

A NEW DEVICE FOR ACHIEVING HOSPITALITY IN ARRAY

One of the most helpful advances in the notation of classification system must be credited to Melvil Dewey: decimal numbers as against integers. The great benefit the use of decimals brings about is the possibility of infinite *hospitality in chain.* This *hospitality* is one of two postulated by S. R. Ranganathan as being necessary for successful translation from the *idea plane* of classification to the *notational plane*. The other necessary hospitality is *in array*.

Hospitality in chain refers to that characteristic of a notation which allows unimpeded *vertical* intercalation of new classes. If our classification scheme begins with two classes, *1* and *2*, any new class will be subordinate to one of them, since by definition *1* + *2* exhausts the universe of knowledge. The first subordinate to *1* might thus be *1½*. Further intercalations in chain (= subordinations) would be *1¼*, etc. Dewey's contribution was in postulating the basic numbers not (as we have given them here) as integers, but as decimals. Subordination below *1* thus is first *15*, intercalation in chain *125, etc.* (read, in order, as *.1*, *.125*, *.15*).

This advantage was originally embodied in the Dewey Decimal Classification, and was adopted by the Universal Decimal Classification and the Colon Classification, but not by the Library of Congress Classification (though it has proven necessary in this last too, though in a somewhat un-overt way and for a somewhat different purpose).

The same notation, however, that brings about (subordinating) hospitality in chain does not necessarily also bring about (coordinating) hospitality in array. Hospitality in array refers to that characteristic of a notation which allows unimpeded *horizontal* intercalation of new classes. If our classification begins by exhausting the universe of knowledge in classes *1* and *2*, any new (*i.e.*, previously unrecognized or unneeded) coordinate main class cannot fall anywhere but outside our numerical base; since, even in a decimal system (base: *0—9*), there can be no number longer than one digit which can be on the same generic/specific level with *1, 2, . . . 9*.

The numerical base we have been using in illustrating the problem is admittedly unrealistically short, but even with a longer one we eventually come upon difficulties. (A short base is used just to make the problem occur at once.) No matter what our base number, even if we use the whole alphabet (as in LC, CC, and Bliss' Bibliographic Classification) there will come the moment when we need more coordinate classes than the notation supplies. Dewey tried to obviate the problem by establishing coordinate classes such as *.1—.8* as specified divisions, with *.9* left over for 'others' —*i.e.*, the unspecified divisions. But if the array *.1—.8* is a systematic one (either canonically, chronologically, or even just alphabetically) it is obvious that the use of subdivisions of *.9* for further coordinates to *.1—.8* destroys the original order. Library of Congress tried to obviate the problem by not assigning all the available numbers when a given genus was broken up into its coordinate species; but once these vacant numbers are exhausted, further intercalation must be accomplished by use of decimals, *etc.** Ranganathan himself, using a decimal system enabling hospitality in chain, has come up with an ingenious device for the enabling of hospitality in array within the same decimal system: the 'octavising' device.

With this device, the array is extendable past its base by the use of the last (or, if the last is needed for some other function, the next last) digit as the equivalent of a musical octave. The series of coordinates thus runs 1, 2, 3, 4, 5, 6, 7, 8, 91, 92, 93, 94, 95, 96, 97, 98, 991, ... If a genus has (say) seventeen coordinate species, they can be accommodated on one coordinate level by *1—991.*

But there seem to me to be heavy objections to the use of this device, either in optically or electronically scanned catalog records:

(*a*) In the case of optically scanned records, there is the great difficulty which would attend its use, not only by the relatively untrained patron, but even by the relatively well-trained librarian. And in the case of electronically scanned records, the desideratum of structural notation as a guide to the computer is lost, or at least would have to be regained by relatively complicated programming.

satisfactory solutions of other requirements would be in conflict.)

4. It should have a quiet electrically driven film transport with variable speeds from creep to 100 ft. in 10 seconds, and the transport must be well designed to prevent scratching of emulsion. Focus must be maintained during slow drive. Switches, film advance, and other moving parts should be designed for relatively silent operation so as not to disturb neighboring readers.
5. The reader should employ an 18" x 24" viewing screen made of transluscent glass. The screen should be adjustable for positions between a vertical incline and horizontal. The image should be viewed directly with both unaided eyes (i.e., no lens or mirrors intercepting). The projection source mechanism must be rotatable through 360° for changes in filming position.
6. The lens, light source and screen should provide an exceedingly sharp evenly-lighted reproduction for normal reading at 15 power enlargement with image resolution of 8 lines/mm. at the edges of the screen and image contrast of between 0.75 to 0.8. It should offer adjustable 20 to 50 foot candles of screen surface illumination and no surface reflection.
7. Areas surrounding the screen within the field of vision should provide a surface with brightness comparable to the screen. This might be accomplished from shaded light sources under the shelf or on the side panels, these luminaries to be adjustable from 20 to 50 foot candles by using a combination of light sourcase and a single rheostat control.
8. Although the general working magnification would probably be in the 12 to 15 power range, the lens system should provide 44 power maximum magnification (providing the possibility of enlarging any part of most texts to twice original size) and 12 power minimum magnification. It should have zoom lens for reaching maximum magnification while holding constant focus.
9. A visible dial should show the footage forward and backward for a full hundred-foot reel of roll film. A set of keys should be available to activate automatic advance to a preselected footage or frame number for film so indexed. A tripping device should automatically stop the advance and trigger an explanatory signal whenever a faulty splice is detected.
10. A dry print-out device should be included with automatic exposure adjustment to reduce waste and delay. The person wishing the copy should himself be able to insert non-photographic paper of several sizes or characteristics (such as paper with preprinted headings or of archival quality).
11. The machine ought to have electrically operated dust hoods, so that when not in use the viewing screen and the lens-blower mounting are enclosed for dust control.
12. Bulbs should be mechanically sturdy and long-lived. The machine should be engineered for simple cleaning of the lens and film holding flats, for easy bulb changing, and for convenient performance of other normal maintenance.

Bibliography

American Standards Association. *American Standard Specification for Microfilm Readers for 16mm. and 35mm. Film on Reels,* PH 5.1–1959.

Ballou, Hubbard W., "What to Look for in a Reading Machine," *ALA Bulletin,* 55:1961, pp. 67–69.

Hawken, William R., *Enlarged Prints from Library Microforms.* Chicago. American Library Association Library Technology Project, 1963. (Esp. "Recommendations" in Chapter 13.)

Walkup, L. E., O. A. Ullrich, J. R. Stock, and J. M. Dugan, The Design of Improved Microimage Readers for Promoting the Utilization of Microimages. *National Microfilm Association Proceedings,* 11: 1962, pp. 283–310.

David C. Weber
Assistant Director of Libraries
Stanford University

Letters to the Editor

Dear Sir: April 14, 1965

I have read with interest the article entitled "Some Clumping Experiments for Associative Retrieval" by A. G. Dale and N. Dale in the January issue. I am presently working with a similar system and wish to point out some difficulties encountered in experimentation, to which the above mentioned article does not give due recognition.

Although the retrieval model presented suggests the output listing of documents are "ordered on some scale of relevancy," this is not reflected in any way by the design of the experiment. To say that "7 of the 10 most relevant documents identified occurred in the first 26 documents" says nothing about the order in which they occur. To present the retrieved documents as a ranked list but not to consider this ranking in evaluation is being inconsistent. The experiment must in some way test the ability of the retrieval strategy to bring the relevant documents to the top of the list and to correctly order these documents.

In addition, we must be cognizant of where in the first 26 documents the 7 relevant ones are located. Before making a statement concerning the ability of the scheme to bring relevant documents to the top of the list, the precise nature of the distribution of the relevant documents in the top 26 positions must be known.

I hope the above comments serve to indicate the inadequacy of present methods of evaluation of a retrieval scheme which produces a ranked output list.

Very truly yours,
ROBERT M. CURTICE
Research Assistant
Center for the Information Sciences
Lehigh University

Dear Sir: April 27, 1965

As one of "those villians" who work on computer applications to documentation problems, I must speak out against a sea of negative attitudes toward computers in the documentation world. I have in mind particularly the question, "Can computers classify better than humans?" There seems to be an irresistible temptation on the part of many to beg this question by calling it meaningless. They say: "How else can you establish how well the computer is doing than by comparing it to human performance?" This in spite of the recent rash of experimental work revealing how much humans disagree with each other.

I am all in favor of a healthy skepticism, as manifested by Isaac Welt's editorial "Information Science-Science Information" in last year's October issue. I am even in favor of conservatism when it comes to day-to-day practice in documentation. But, I believe this attitude out of place when extended to the prospects of research and development. It is true that many computerites have overstepped themselves in wild-eyed enthusiasm, but this is no signal for reactionaries to move into City Hall. Let the people who are dealing with today's information problems and the people who look at the possibilities of tomorrow understand each other and not get their feet tangled up in each other's interests.

On the issue of "computers vs. humans in classification," it is a truism to say that we have hardly begun to find out what computers can do. Some dubiously conceived experiments which have come to my recent attention have caused everything from sighs of relief to cackling in some circles because they seem to prove "once and for all" that computers are unsuitable for the "intellectual task" (after all!) of classification. Admittedly even under the best of conditions classifications based on computerized statistical analysis aren't yet as good as what people do in the best libraries. However—and this is the key point I want to make —if today we had to put up with Edison's first version of the electric light, we would all go back to kerosene lamps. So let's not knock progress merely because it's not quite here yet.

Sincerely yours,
LAUREN B. DOYLE
Language Processing
and Retrieval Staff
System Development Corporation

Dear Sir: April 9, 1965

In view of the present trends in office copying machines, bound journal volume size, and information retrieval and dissemination problems, I have a suggestion to make. It is that, from the beginning of the year 1966, all journals and technical books should be printed with a ¼" outer margin and a 1½" inner margin. It will not thereafter be necessary to break the back of library books in order to obtain a copy of a graph or other interesting data. I know this will require a break with tradition, but then, so was the emancipation proclamation.

Yours sincerely,
RICHARD J. MCCLURE
Senior Research Physicist
Bell & Howell Research Center

LITERATURE NOTES

BURTON E. LAMKIN, Editor

Contributors of abstracts from readers and suggestions of books and articles for review or inclusion in this bibliography will be welcome and are actively solicited. Volunteer abstractors and reviewers are needed, as well as people with special linguistic talents. All copies of reprints, reports, and correspondence for this section should be addressed to Mr. Burton E. Lamkin, 1649 Fairorchard Ave., San Jose, Calif., 95125. In Order to increase its coverage, this section will include some abstracts copied or adapted from other publications. The American Documentation Institute is not able to supply copies of the publications abstracted or cited.

Abstractors

The following persons have prepared some of the abstracts for this issue. Their names are listed here to express the editor's thanks and appreciation, as well as to explain the initials that were used to sign the abstracts.

C.L.B.	Curtis L. Brown	C.B.L.	Charles B. Leidenfrost
M.H.G.	Maria H. Grimes	F.S.	Frantz Spachner
H.W.B.	Henry W. Brann	T.A.T.	Teresa A. Townsend
S.A.	Susan Artandi	D.R.M.	Dorothy R. Meller
R.R.F.	Bob Freeman	L.D.	Laszlo Dosa

Contents

Starting with the January 1965 issue, separate copies of this section of *American Documentation* may be obtained as reprints on a subscription basis for $4.00 per year. Requests should be directed to ADI, 2000 P Street N.W., Washington, D. C., 20036. It is hoped that in the very near future it will be possible to provide Literature Notes as a publication separate from *American Documentation* and with increased coverage. The availability of separate reprint sections is a start toward that objective.

Index of Authors and Corporate Sources Referred to in this Issue

(The numbers refer to the terminal digits of the abstract serial numbers in this issue. The letter R refers to Reviews.)

Journal Abbreviations

Many of the journal titles are represented by their corresponding CODEN notation as defined in the two publications, *Coden for Periodical Titles,* American Society for Testing and Materials, Phila., Pa., 1963, ASTM Special Technical Publication No. 329, and *Coden for Periodical Titles, Supplement 1,* 1964, ASTM Special Technical Publication No. 329 — 31. The CODENS are being used in place of the full titles primarily to reduce typing and typesetting work, and to provide a more compact representation of the citation. The following table will help translate the CODENS for this issue.

ALBL	ALA Bulletin
AMDO	American Documentation
ARHE	Arthritis and Rheumatism
ASLP	Aslib Proceedings
BABL	Babel
BATR	Battelle Technical Review
BEHS	Behavioral Sciences
BHOR	Business Horizons
BLRC	Bell Laboratories Record
BMLA	Bull. Medical Lib. Assoc.
BRNS	Barron's
BUAT	Business Automation
CACM	ACM Communications
CENE	Chemical and Engineering News
CENG	Control Engineering
CMPG	Computer Journal
CPGR	ACM Computing Reviews
DKMN	Dokumentation
DTMN	Datamation
ERGO	Ergonomics
FEPR	Federation Proceedings
FRON	Frontier
IBMJ	IBM Journal
IEAI	IEEE Trans. on Applications & Industry
IEEI	IEEE Trans. on Ind. Electr. & Control Instr.
IEEP	Proc. IEEE
IETT	IEEE Trans. on Information Theory
IEEW	IEEE Trans. on Eng. Writing & Speech
IPAL	Inland Printer and American Lithographer
IREL	IEEE Trans. on Electronic Computers
IRTE	IEEE Trans. on Eng. Management
ISCT	Int'l. Science & Technology
ISSR	Information Storage & Retrieval
JACM	ACM Journal
JCHD	J. Chemical Documentation
JDOC	J. Documentation
LANS	Language and Speech
LCIB	Library of Congress Information Bull.
LIBJ	Library Journal
LIBL	Library Literature
LIBQ	Library Quarterly
LIBT	Library Trends
LRTS	Library Resources & Technical Services
MTRS	Materials Research & Standards
NADO	Nachrichten für Dokumentation
NAMN	National Micro-News
NMRV	Navy Management Review
NWSC	New Scientist
OPRE	Operations Research
PMOS	Perceptual and Motor Skills
PRMG	Printing Magazine/National Lithogr.
REDE	Research/Development
REDO	Revue Internationale de la Documentation

REPM	Reproduction Methods for Business & Ind.	SHBN	Stechert-Hafner Book News	TAPP	Tappi
RESM	Research Management	SPLB	Special Libraries	UNDL	UNESCO Bull. for Libraries
RPRV	Reproduction Review	SSTM	Systems	WLDU	Wilson Library Bull.
SCIE	Science	STNW	Sci-Tech News		
		TAGU	Trans. of the Amer. Geophysical Union		

Conference Proceedings and Collections of Papers

7/65–1 **Proceedings of the 13th Annual Meeting and Convention.** Philadelphia, Pa. April 28-30, 1964. Vernon D. Tate, Ed. National Microfilm Assoc., Annapolis, Maryland. 1964.

Published Reference Tools

7/65–2 **Minnesota Standards on Microfilming Public Records.** Wayne A. VanderVort. NAMN, no 73, December 1964, pp. 119–125.

7/65–3 **Notes on Standardization Literature.** L. L. Lawrence. NAMN, no. 72, October 1964, pp. 52–56.

7/65–4 **A Thesaurus of Civil Defense Descriptors: Appendix E of Final Report (Volume II) Civil Defense Information Systems Analysis.** January 1965. W. T. Herzog and J. E. Jenkins. The Research Triangle Institute, Operations Research and Economics Division, Durham, N.C. Prepared for: Department of the Army, Office of the Secretary of the Army, Office of Civil Defense, Research Directorate. Contract OCD-PS-64-56. OCD Subtask 4631B. RTI Number OU-158-7.

A listing of approximately 600 words or phrases (descriptors) is provided for use in indexing and retrieving civil defense research information. Inclusion notes, "Use," and "Also see' references are provided as well as general directions for applying descriptors to civil defense information. The selection has been modeled on the system of the Defense Documentation Center (DDC). About 70 of the 600 descriptors are not used by DDC but were chosen from the Civil Defense Matrix and from special fields of interest. This 600-word core can be supplemented with specific terms in the subject areas where the quantity of documents warrants more intensive indexing. From previous library experience, it is estimated that 600 is the minimum number needed for a collection of 1000 documents. (authors) (see 7/65–62)

7/65–5 **Guide to Microforms in Print.** 1965. Albert J. Diaz, Ed. Microcard Editions, Inc., 901 26th Street, N.W., Washington, D.C.; price $4.00.

7/65–6 **Handbook of Russian and Soviet Bibliographies: A listing of general and specialized bibliographies and reference works of Russia and the Soviet Union.** I. Kirpicheva, translated into German by Regina Stein and Heinz Hohne, Leipzig, Verl. f. Buch-und Bibliothekswesen, 1962, 225 pp. Reviewed in *Zentralblatt fur Bibliothekswesen*, (Central Newspaper of Library Science), no. 2, 1962, pp. 100–102. In German.

The first part lists second-level bibliographies of books, periodicals, magazine and newspaper articles and reviews, and includes a discussion of bibliographies published prior to the October Revolution. The second part contains important bibliographies and reference works arranged by individual scientific fields, each chapter beginning with bibliographies of current information and concluding with a list of literature from the 18th to the 20th century. The last part contains a list of Russian abbreviations and an alphabetical listing of authors and titles. Included in the Russian original but omitted in this second German edition is a list of Russian and Soviet general encyclopedias. (M.H.G.)

7/65–7 **Information Retrieval—a Solution to Todays "Literature Explosion."** Winter, 1964. Peter M. Nobbs. *Trend—Activities Pulp Paper Res. Inst. Can.* no. 4, pp.12–16.

A general introduction to concept-coordination indexing is followed by the announcement and description of the "Thesaurus of Pulp and Paper Terms," published by and available from the Pulp & Paper Research Institute of Canada (Montreal). This was developed, as a supplement to the *EJC Thesaurus of Engineering Terms*, through the cooperative efforts of PPRIC, The Inst. of Paper Chemistry (Appleton, Wis.), and several Canadian and U.S. paper companies. (C.L.B.)

7/65–8 **Guide to Libraries and Literature Depositories in the Capital City of Berlin.** 1963. Leipzig, Verl, f. Bunch-und Bibliotheskswesen, 254 pp. Reviewed in *Zentralblatt fur Bibliothekswesen* (Central Newspaper of Library Science), no. 2, 1964, pp. 96–99. In German.

This new guide is limited to East Berlin and adjacent areas and lists 550 libraries and information and documentation centers. It reports seven scientific general libraries with a total of 6,694,714 volumes and 426 specialized scientific libraries and documentation and information centers with a total of 6,116,947 titles. The libraries listed have increased their total holdings by almost 3.5 million volumes since 1950. In addition to providing essential data concerning holdings in individual libraries, the guide provides information about circulation facilities, cataloging, and information and consultation services. Titles of periodical publications by and about each installation also are supplied. The last chapter includes an alphabetical index of libraries, a subject index, an index documentation and information centers, an index of regional libraries and an index of school and university libraries. (M. H. G.)

7/65–9 **Library Information Service. List of Centers.** 1963. M. Kociecka, Warsaw. National Library. 66 pp. Reviewed in *Bibliotekarz* (*Librarian*), No. 2/3, 1964, p. 90. In Polish.

This publication contains a list of library information centers compiled by the Institute of Scientific Information of the National Library. One hundred sixty seven libraries of various types are included. In addition, many special or technical libraries are included. The list gives the name of the center, its specialization, the type of catalog used, and its methods of distributing information to patrons. (T. A. T.)

7/65–10 **Czech Periodicals in the Jagiellonian Library Collection.** J. Kowalik, *Biuletyn Biblioteki Jagiellonskiei* (*Jagiellonian Library Bulletin*), no. 1, 1963, pp. 5–12. In Polish.

The library divides its Czech collection into the following categories:

1. Books published to 1868.
2. 1868–1905 (during this period the Jagiellonian Library was directed by Karol Estreicher, who strongly supported the work on the Czech periodical collection).
3. First World War period.
4. Period between First World War and Second.
5. 1945 to the present.

The article lists some interesting facts on the methods used to acquire the collections. During the first period, most of the acquisitions came as gifts from the Scientific Society of Krakow (1816–1855), others were gifts from individuals, and some came in exchanges with Prague University. During the second period, gifts

remained the main source of acquisition. Many Czech institutions and individuals sent large collections to the library. When Karol Estreicher left the Library, the rate of growth of the Czech collections slowed down, and World War I further reduced the amount of new material being added. From 1930 to 1945, the main source of acquisition was international exchange. Since 1945, the Czech periodicals are acquired mainly by subscriptions and exchange. (T. A. T.)

7/65–11 **Guide to Unpublished Polish Bibliographies,** B. Eychler, *Biuletyn Instytutu Bibliograficznego* (*Bulletin of the Bibliographical Institute*), no. 1, pp. 1–139. In Polish.

In the short introduction the author explains that his primary purpose is to make available to all libraries, institutions, and individuals the results of the central registration of all bibliographical research that is being planned, worked on or finished. He also explains what the listing includes and how it was prepared. The data were collected by the Bibliographical Institute of the National Library. The bibliographical information is broken down into several main categories: bibliographical information is broken down into several main categories: bibliography of bibliographies; general; by subject matter; regional and local bibliography; ethnic on various institutions of higher learning, libraries, publishing houses and persons. The guide also includes bibliographies of periodicals and other types of published matter such as calendars, iconography, maps, memoirs, etc. An index is provided of subjects, authors and institutions conducting bibliographical research. (T. A. T.)

7/65–12 **Soviet Libraries and Their Information Services,** G. Pomassl, *Zentralblatt fur Bibliothekswesen* (*Central Newspaper of Library Science*), no. 3, 1964, pp. 129–148. In German.

The article is a report of a tour of Soviet libraries. The purpose of the study was to obtain helpful data for establishing an effective information and documentation system in East Germany. Soviet scientific libraries are important information centers for domestic and foreign source material. They select, process, and disseminate specialized information according to the overall needs of the users. Systematic coordination of the information activities of individual institutions is established on a local and regional basis. The Soviet libraries also actively support developments in the theory and practice of information science. Bibliographical services are considered to be a very important component of the documentation work. Due to effective coordination characteristic traits of the bibliographical activity of each library are readily discernible. Provisions are made for graduate-level training of information and documentation specialists in general and specialized scientific libraries. Bulletins and flash sheets are issued regularly for users of the documentation and information services. (M. H. G.)

7/65–13 **Geoscience Abstracts Index: Subject (UNIDEK) and Author,** Vol. 6, No. 12, Pt. 2 of *GeoScience Abstracts*, December 1964, Washington, American Geological Institute.

The subject index is a computer-compiled (IBM 1401) systematic index based on the Universal Decimal Classification of the Federation Internacionale de Documentation. It is supplemented by a systematic listing of the major index headings and an alphabetic listing of all significant words from all headings which were used throughout the volume of GeoScience Abstracts. The systematic and alphabetic listings are used as guides to the UNIDEK, which contains complete titles and corresponding abstract numbers. (R.R.F.)

7/65–14 **A Thesaurus of Rheumatology.** M. J. Ruhl and L. Sokoloff. ARHE, vol. 8 no. 1, part II. February 1965. pp. 97–182.

A novel thesaurus, cataloguing more than 1100 terms, has been constructed. It has been designed to serve four purposes: (1) to assist in adapting the subject headings employed by MEDLARS for indexing the more numerous rheumatologic terms accurately in connection with the preparation of the *Index Medicus*, the *Index of Rheumatology*, and also specific demand searches; (2) to guide the *Index of Rheumatology* and other literature analysis programs as to the specific subjects to be included in their coverage; (3) to serve as a rheumatologic vocabulary, systematizing diverse eponyms, synonyms and usages; and (4) to serve as a first reference source for literature search. Reprints may be obtained from the American Rheumatism Association at cost.

7/65–15 **Thesaurus of Pulp and Paper Terms; a list of keywords with their cross-relatinoships for use in indexing and retrieving information on pulp and paper technology.** First edition, March 1965. 96 pp. Available from Pulp & Paper Research Inst. of Canada, 570 St. John's Rd., Pointe Claire, Quebec, price $10.00. (See 7/65–7) (C.L.B.)

Characteristics of Information Resources

7/65–16 **Aid-to-Indexing Forms. Progress Report No. 2.** December 1964. Pauline Atherton. Report AIP/DRP–64–2 of the American Institute of Physics. New York, N.Y. 14 pp. + appendix.

This is a report of the experience gained in using aid-to-indexing forms received from authors of over 2,000 papers which appeared in six physics research journals. Examples of executed forms, some revised since AIP/DRP 63–2, are included. The response from authors has been very good. The information obtained is useful for amplifying titles, evaluating abstracts, indicating type of research reported, and choosing subject index entries. The use of the forms by abstract-journal editors and for the experimental development of better indexes is also reported. (author)

7/65–17 **Organization of a National Scientific and Technical Information Center.** October 7, 1964. Wilfred James Mayo-Wells, Washington, D.C. (AD–445 078.) 120 pp.

Contents: Dissemination of Scientific and Technical Information; Selection of the Optimum Plan to Achieve Objectives; Selection of the Criteria by which the Information Disseminated by NIAC should be Judged; The Functions and Operations of a Scientific and Technical Information Center. (author)

7/65–18 **Citations of Physics Literature.** Experimental Issue. 1964. Ben-Ami Lipetz and Charlotte F. Dziurkiewicz. Report AIP/DRP-C1–1964 of the Documentation Research Project of the American Institute of Physics. New York, N.Y. 105 pp.

This index is purely experimental. It was prepared as part of a program to evaluate the potential usefulness of the citation index as a reference aid for physicists. It is hoped that it will prove useful in helping American physicists to find relevant Russian work. It lists all papers in a limited body of source literature which cite specific papers in *Journal of Applied Physics or Physical Review*. The source literature consists of all issues published through June 1963 of four of the journals in the Soviet Physics translation series. These journals are: *Soviet Physics—Crystallography; Soviet Physics—JETP; Soviet Physics—Solid State;* and *Soviet Physics—Technical Physics.*

Access to this index is through volume and page numbers of pertinent papers published in the cited journals (*Journal of Applied Physics or Physical Review*). Starting with precise volume-and-page identifications of known papers, the reader can determine whether and where these papers were referred to in the Soviet Physics journals. (authors)

7/65–19 **Hurdling the Language Barrier.** Robert T. Beyer. *Physics Today*, vol. 18, no. 1. January 1965. pp. 46, 48, 50.

English is still the major language of physics literature, accounting for 76% of 3000 abstracts sampled from *Physics Abstracts* and for 63% of 450 abstracts (in theoretical, elementary particle, and magnetic physics) sampled from the Russian *Referativnyi Zhurnal*. Russian was second, viz., 14% of the *Physics Abstracts* sample (of which 12% was adequately covered by the full-journal translations of the "Soviet Physics" series, and 2% was not covered in English), and 24% of the *Referativnyi Zhurnal* sample. (C.L.B.)

7/65–20 **Domestic and International Interlibrary Loans.** Anon. *Modszertani Korlevel (Circular on Methodology)*, no. 15, 1964. p. 10. In Hungarian.

Merits of Hungarian interlibrary loan services, whose main coordinator is the National Szechenyi Library in Budapest, are discussed. Emphasis is placed on the savings in foreign exchange currency made possible through international loans. Instructions are given for processing interlibrary loan requests; and the forms used by the libraries for this purpose are illustrated. (L.D.)

7/65–21 **Service to Industry—The Facilities at RAPRA.** Anon. *Packaging News* (London), vol. 12, no. 2, February 1965. pp. 39–41.

The recent merger between Brit. Plastics Federation and Rubber & Plastics Research Assoc. (RAPRA) resulted in a merger of library and IR facilities as well. The history and organization (library, research, & technology divisions) of RAPRA are outlined. Publications include *Rubber Abstracts, Plastics Abstracts, RAPRA Bull., Foreign Rubber Patents, Material Evaluation Reports*, and irregular research reports, technical reviews, etc. Member companies can avail themselves of translation and speedy (usually 24-hr) question-answer search services. Library holdings comprise 15,000 books, 500 journal titles, 90,000 technical pamphlets, miscellaneous materials, and catalog-index files numbering over 1 million cards. (C.L.B.)

7/65–22 **Contradictions Between Research and Scientific Information.** C. Rozsa. *Magyar Konyvszemle (Hungarian Book Review)*, no. 1, 1964, pp. 33–45. Published in Hungary in French.

The following contradictions are investigated in this article: definite research requirements as opposed to a mass of unspecified, loosely identified knowledge; the discrepancy between the large amount of publications and documents and the modest amount of useful information in them; the exponential growth of scientific publications and the exponential decay with time of the value of the facts this literature contains; the contradiction between the diversity of contents and concentration on selected topics in special literature; demands for more specific scientific data by the users, and the increasingly complex and universal structure of science; the tendency of scientific information services to expand and comparatively limited use made of them; scientific information, bridging the gap between the researcher and the source of information, and at the same time widening it; the lack of a universal system of science classification and the need of a classification system for documentation; the contradiction between the autonomy of research centers and scientific information centers; research and information competing for qualified specialists; increase in the information activities and the relatively slow development of its theoretical foundations. (Doc Inc)

7/65–23 **SIE Seeks to Expand Research Coverage.** CENE, vol. 43, no. 1, January 4, 1965. pp. 28–29.

Briefly describes the Science Information Exchange of the Smithsonian Institution.

7/65–24 **Toward a National Chemical Information Network.** Walter M. Carlson. JCHD, vol. 5, no. 1, February 1965. pp. 1–3.

Analyzes the distribution of information basic concepts as: purpose—current awareness and task-oriented search; scope—scientific and technical information; orientation—general interest, mission or project interest, and management interests. (BATR)

7/65–25 **Private and Public Communications in Physics.** Michael J. Moravesik. *Physics Today*, vol. 18, March 1965. pp. 23–26.

Gives suggestions for the utilization and improvement of printed communications, refereeing systems, preprints, and private communications. (BATR)

7/65–26 **Information Services in the Field of Economics.** S. Zieminski. *Aktualne Problemy Informacji i Dokumentacji (Current Problems of Information and Documentation)*, no. 1, 1964. pp. 14–17. In Polish.

Informational needs in the field of economics and specific characteristics of such information are described. The best forms and methods of the information service needed to meet the particular requirements of the national economy and administration are analyzed. Various institutes and information centers are active in this field, and the services they offer are listed. Suggestions are made regarding the coordination of work and the division of responsibilities between the Central Institute of Technical Economic Information (CIINTE), and the information centers of various branches of industry. (Doc Inc)

7/65–27 **The Literature Analysis Project of the American Rheumatism Association.** Mary Jane Ruhl and Leon Sokoloff. ARHE, vol. 7, no. 6, December 1964. pp. 615–622.

A Literature Analysis Project of the American Rheumatism Association has taken three steps to improve the dissemination and retrieval of information on rheumatic diseases: (1) The domain of rheumatology and the limits to which basic research literature should be included in its bibliographic services were delineated by a questionnaire survey. (2) A novel *Thesaurus of Rheumatology* was prepared to improve the indexing of publications and to codify the vocabulary; (3) A computer-driven, semimonthly bibliography of current publications, the *Index of Rheumatology*, was developed with the cooperation of the National Library of Medicine. This bibliography is anticipated to yield approximately 6000 indexed citations per year. The proportions of relevant papers missed in the 2,299 periodicals covered and the irrelevant ones included are estimated to be below 20 and 10 percent of the total respectively. (authors)

7/65–28 **The Structure of the Information and Documentation System in East Germany.** A. Derfl. *Knihovnik (The Librarian)*, no. 3, 1964. pp. 74–80. In Czech.

The article describes information and documentation services in East Germany as seen by a Czechoslovak specialist. Until recently, the Documentation Institute of the German Academy of Sciences directed and administered a network of centers dealing primarily with scientific, technical and public health matters. Close ties between research and production formed the basis for practical use of new data and for planning. Instructors responsible for the methodical direction of the centers coordinated their activities. A working branch of the Documentation Institute was established in Leipzig. There were other centers for information and documentation, some of which carried out their activities independently, including publicity, international cooperation, etc.; others were tied closely to the activities of the Documentation Institute. The organization of the information service thus was similar to that in Czechoslovakia except for decentralization in documentation service. No central card catalog was established, and cooperation with other agencies was poor.

The technical library system in East Germany was established independently of documentation and information centers, and under separate administration. Forms and methods of library work have been systematically adapted to the need of scientists and engineers. The Information Center of the German Library (Deutsche Bucherei) in Leipzig prepares bibliographies on various subjects on request. Other services include a card index of abstracts, patent files, and collection of industry specifications and standards. Request bibliographies are also prepared by the Scientific-Technical Division of the Library of the Dresden Technical Academy. In August 1963, the State Planning Commission with its Central Institute for Information and Documentation was made responsible for all information and documentation activities. While the Documentation Institute limits itself henceforth to activities sponsored by the Academy of Sciences, the Central Institute for Information and Documentation controls, coordinates, and administers information and documentation services; directs all theoretical work; trains information experts; publishes pertinent literature; represents East Germany on the international

level; promotes research aimed at mechanizing and automating information services; and provides a central information service for leading governmental offices. (M.H.G.)

7/65–29 **Information Services by Medical Libraries.** Anon. *Az Orvosi Konyvtaros (The Medical Librarian),* no. 3/4, 1963. pp. 79–98. In Hungarian.

Since the world output of medical periodicals exceeds 6,000 publications, the survey of literature would take approximately 30% of a scientist's time; therefore, the librarian's task is not only to collect but also to disseminate pertinent information. This kind of information service requires special skills and it should include: (1) acquisition of the right material (using domestic and foreign reviews); (2) publishing contents lists of periodicals; (3) publication of reference materials (by author or title, field of interest, and subject matter); (4) preparation of research or subject bibliographies; (5) other services, e.g., translations, photocopies, microfilms. However, not all the Hungarian medical libraries can provide these services. Only a few larger libraries are in a better position like the Documentation Center of Medical Science which publishes a *Hungarian Medical Bibliography.* Specific services of some provincial medical libraries are described. (C.B.L.)

7/65–30 **Organization of Information and Documentation Services in East Germany.** A. Derfl. *Technicka Knihovna (Technical Library),* no. 3, 1964. pp. 68–71. InCzech.

Recent changes in the organization of East German industry had considerable effect on the structure of the information and documentation services, too. Until now, the Institut fuer Dokumentation of the German Academy of Sciences was the coordinating agency for individual centers, including those in the factories. It also published important bibliographical periodicals. Under the new system, the recently established Central Documentation and Information Institute is subordinate to the State Planning Commission. It is responsible for establishing a uniform and effective information and documentation system and wide coordination and organization of these services. It cooperates with the Council for Mutual Economic Aid to promote specialization and division of labor, and makes arrangements with scientific libraries and publishing companies to standardize information and documentation services. Central administration organs, central research institutes, scientific institutions, and individual enterprises are responsible for information and documentation activities within their own jurisdictions. Their responsibilities are described. Institute fuer Dokumentation is to limit its activities to requirements of the German Academy of Sciences. It will continue to publish the basic bibliographical periodicals *Chemisches und Technisches Zentralblatt (Central Chemical and Technical News)* and *Kernforschung und Kerntechnik (Nuclear Research and Technology).* (M.H.G.)

7/65–31 **Acquisition Center for Libraries.** J. Irmler. *Citatel, (The Reader),* no. 5, 1964. pp. 170–171. In Slovak.

The author advocates the establishment of a special center to provide planned, systematic acquisitions for public and special libraries. It should be part of the national enterprise Slovak Book (Slovenska Kniha) and would coordinate with the activities of administrative and research institutions of the Slovak library network. Emphasis would be placed on Czech technical and scientific books. It would also supply all libraries with forms and equipment, and in addition, the proposed center would collect obsolete literature and store literature for special group projects. A large number of copies of a single title generally is required for such projects; upon completion, the copies would be returned to the center to be distributed elsewhere or sold. It also would supply the libraries with selected book packages. There are 400–500 titles which every person is expected to read, but which are not available in libraries in sufficient numbers. (M.H.G.)

7/65–32 **Notes on the Organizational Forms of Technical Information (Analysis of the Coverage of Selected Topics on Transportation and Communication in Special Literature).** G. Hovanyi. *Orzsagos Muszaki Konyvtar es Dokumentacios Kozpont Evkonyv—1962 (Hungarian Central Technical Library and Documentation Center, 1962 Yearbook),* Budapest, 1963. pp. 172–213. In Hungarian.

Among other things, the subject monitoring service of the Hungarian Central Technical Library and Documentation Center is scranning about 1,600 different periodicals for 141 selected topics from the fields of transportation and communication. The author analyzed the material collected during 1962 and evaluated it according to Bradford's law of scattering. The requirements of the requestors and users of these services are evaluated too, since they affect the extent of scattering. The analysis disclosed that in the special literature on transportation and communication, 48% of the articles pertaining to a specific branch of this field appeared elsewhere than in the periodicals of that particular branch. Moreover, 38% of articles dealing with transportation and communication problems in general were published in periodicals which are not directly concerned with transportation and communication. The effects of the definition of the areas of interest, the efficiency of the coverage of a specific field, the exhaustiveness of the reference sections of technical periodicals, and other related aspects are discussed. It was concluded that an analysis of scattering made by a different method than Bradford's yielded results identical. The results also proved the feasibility and practicability of establishing a central institution for scanning and evaluating special and technical literature of all branches of science. (Doc Inc)

7/65–33 **Certain Problems of Inter-Library Loans.** M. Wigluszowa. *Bibliotekarz* (Librarian), no. 1, 1964. pp. 12–15. In Polish.

This article discusses ways of improving systems of inter-library loans, which have become very popular in recent years in connection with the increase of scientific research in Poland. Certain specialized libraries are preparing catalogs of pertinent foreign publications. These are kept up to date and include information concerning subscriptions and recent changes. The author feels that such catalogs are most helpful and should be popularized, since they often give access to lesser known publications. He believes that articles in most popular periodicals should be copied and duplicated to avoid monopoly of the original by one person. A deficiency in the supply of certain foreign publications requires ordering of microfilm copies of individual articles from abroad, which is time-consuming and expensive. The majority of readers prefer to use an original publication rather than microfilm or photocopy. It seems that copies from copying machines, such as the American made Thermo Copying Machine at the Central Agricultural Library in Warsaw, or the Polish made Xerograf, are more economical and easier to use than microfilm. (T.A.T.)

7/65–34 **Technical Information Specialists Help to Disseminate Scientific-Technical Information in Factories.** J. Podzimek. *Podnikova Organizace (Factory Organization),* no. 2, 1964. pp. 70–72. In Czech.

It is a common practice in the Soviet Union to have technical information specialists in factories and plants, where they function as an extended arm of the scientific-technical information centers. This practice finds an ever-increasing application in Czechoslovakia. The duties and responsibilities of the technical information specialists are described. Among other things, they are responsible for the dissemination of technical information among the engineers and workers of the factory or plant, organization of study circles and seminars, detection and removal of bottlenecks in the production process, organization of visits to study production methods of other enterprises, publicizing and promoting inventions, etc. Working relations between technical information specialists and information documentation centers are also discussed. (Doc Inc)

7/65–35 **Should Microfilms Replace the Original Copy in Interlibrary Loan?** L. Penzold. *Zentralblatt fuer Bibliothekswesen (Central Journal of Library Science),* no. 9, 1964. pp. 526–528. In German.

The author discusses whether it is more practical and economical to use original copies of microfilm for interlibrary loans.

Costs of processing, packaging, and shipping of originals vs. microfilm in East Germany are compared. The author concludes that microfilm is preferable to original copies for the following reasons: the originals remain in house and are available to the users except for the short time of microfilming; shipping costs of microfilm are considerably smaller; amount of labor for packaging microfilm is smaller; books are not exposed to potential damage during shipping; and the user can retain the microfilm. (Doc Inc)

7/65–36 **More On Central Cataloging.** Stenfan. *Ctenar (The Reader)*, no. 2, 1964, p. 71. In Czech.

The author analyzes the first year of experience with central book cataloging of Slovak books in Slovakia. The system was established to promote a systematic catalog, manpower saving, and uniform cataloging rates. Its success is retarded by the inferior quality of paper, technical discrepancies on book slips, different formats of catalog cards, nonuniform type of print, and occasionally by faulty interpretation of cataloging rules. Brief annotations of book content are desirable. The author endorses the use of centralized cataloging, not only for Slovakia but throughout Czechoslovakia. (M.H.G.)

7/65–37 **A Symposium on the Publication of Mathematical Literature.** I. E. Block and R. F. Drenick, Eds. *SIAM Review*, vol. 6, no. 4, October 1964. pp. 431–454.

Sponsored by the Society for Industrial and Applied Mathematics in New York, October 1963. The participants were L. Bers, R. P. Boas, T. Minder, C. Stoll, and H. A. Wooster (Moderator). (BATR)

7/65–38 **Scientific Information Exchange in Psychology.** William D. Garvey and Belver C. Griffith. SCIE, vol. 146, December 25, 1964. pp. 1655–1659.

Gives results of dissemination study performed by the Project on Scientific Information Exchange in Psychology, American Psychological Association, Washington, D.C. (BATR)

7/65–39 **Cooperation Between Public and Technical Libraries,** J. Petraskova, *Statni Technicka Knihovna v Praze. Metodicky Letak. (National Technical Library in Prague. Methodical Flashsheet)*, no. 38, 1963, pp. 1–7. In Czech.

Public libraries are expected to make a contribution to the dissemination of technical knoweldge. The activities and responsibilities of public and technical libraries in their particular fields are reviewed. It is concluded that there is a need to agree upon a coordinated approach to the problem of dissemination of technical knowledge and to work out a plan which would reflect the main objectives in this field. The author makes various suggestions how public libraries could help technical libraries and vice versa. (Doc Inc)

7/65–40 **Organization of Scientific-Technical Information.** A translation of an article by V. S. Malov in the Russian-language periodical *Nauchno-tekhnicheskaya informatsiya (Scientific-Technical Information)*, no. 6, Moscow, 1964. pp. 3–5. *Foreign Developments in Machine Translation and Information Processing*, no. 163. U.S. Department of Commerce, Clearinghouse for Federal Scientific and Technical Information, JPRS: 27,716, TT: 64–51881; price: $2.00.

In our country technical progress is directly related to the study and dissemination of the achievements of Soviet and foreign science and technology and of the progressive experience of enterprises. "In our century of tempestuous scientific-technical progress," said Comrade N. S. Khrushchev, "society cannot develop without a systematic and all-round use of scientific achievements." In this connection, the organization of scientific-technical information aimed at supplying scientists, engineers and workers with information on scientific and technical achievements is particularly important. Basically, scientific-technical information is, at the present time, an organic part of the complex of scientific research, of designing and construction work. (author)

7/65–41 **Measures for the Further Improvement of the System of Scientific-Technical Information in the Country.** A translation of an article by N. B. Arutyunov in the Russian-language periodical *Nauchno-tekhnicheskaya informatsiya (Scientific-Technical Information)*, no. 6, Moscow, 1964. pp. 6–12. *Foreign Developments in Machine Translation and Information Processing*, no. 163. U.S. Department of Commerce, Clearinghouse for Federal Scientific and Technical Information, JPRS: 27,716, TT: 64–51881; price: $2.00.

Manual Card Systems

7/65–42 **Random Retrieval of Micro-Image Parts.** 1964. Lester G. Stine. NMA Proc. (see 7/65–1)

In the past, considerable emphasis has been placed on roll film image retrieval techniques. Selection methods for film card systems; aperture or microfiche, have not developed much past the manual approach. More recently, development of random card selection systems appears appropriate for the handling of unitized film card records. The evolution of the "Needlesort Concept," automatic card selection and retrieval procedures, and the increased potential of micro-image cards through random file techniques will be discussed. (author)

7/65–43 **The Use of Edge Notch Cards in the National Museum.** M. Sour. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers)*, no. 2, 1963. pp. 48–56. In Czech.

The use of edge notch cards for the reference file of the National Museum in Prague, Czechoslovakia, is described. The references are arranged according to language, date, and type of source (personalia, correspondence, and manuscripts). An example is cited of a project for using edge notch cards in the filing of documentary material on the science of geodesy and geology. (Doc. Inc.)

7/65–44 **Edge-Notched Punched Cards in Libraries and Technical-Scientific Information Centers.** J. Bukovsky. *Metodicke Letaky (Methodological Pamphlet)*, no. 36, 1964. pp. 1–28. In Czech.

The article describes format for the available edge-notched punch card, various methods of coding (numerical, alphabetic, numerical-selector codes, etc.), punching techniques, and support equipment. Examples of various codes for recording bibliographic data, classification numbers, etc. are given. Application of the punch cards in library operations for indexing periodicals, books, patents, research reports, as well as for maintaining files of readers, request items, etc. are explained. (Doc. Inc.)

7/65–45 **Small Sorting Systems in Documentation Work.** R. Palatova. *Mechanizace a Automatizace Administrativy (Mechanization and Automation of Administration)*, no. 12, 1964. pp. 294-295. In Slovak.

Potential improvements in the mechanization of information processing are investigated. The edge-notch and edge-punch card sorting system analizator and the aspect card sorting systems designed by Palatova and Helbich are described. Comparative studies were made on the time required to store and retrieve information by entering and classifying data on regular file cards and by the three sorting systems. It was established that the aspect card system is the least time consuming of the three systems both from the storage and retrieval point of view. In addition, it requires only a fraction of the time needed to store and retrieve information by the regular classification and filing method. (Doc. Inc.)

7/65–46 **The Use of Peek-A-Boo Cards for the Microfilm Store of Articles.** J. Toman. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers)*, no. 2, 1963. pp. 43–47. In Czech.

Problems of filing microfilms according to subject matter and the suitability of a system of classification by categories, which should facilitate improved information retrieval, are discussed. In dealing with microfilms the peek-a-boo system was chosen in preference to the Erwool system and edge notch cards. Procedures for using the chosen system are described. This system provides a tool for a deeper analysis of articles, and each inquiry can be searched from various viewpoints. (Doc. Inc.)

Reports of Facilities in Operation—Libraries

7/65–47 **Slavic Collection Retrieval System Manual.** December 1964. Mary Wightman, Clay Swinburn, and A. T. ten Broeke. Report LRC–64–TS–2 of the Linguistics Research Center. University of Texas, Austin, Texas. NSF grant GN–308.

The classification scheme used to index a magnetic tape file of Slavic materials contained in the Main Library of The University of Texas is specified. The indexed bibliography is used as a basis for producing selective listings using a coordinate retrieval program developed at the Linguistics Research Center. Instructions for submitting requests for selective bibliographic listings are given. (authors)

7/65–48 **NOL Retrieval System for the 7090.** October 1, 1964. James J. Crockett. Naval Ordnance Lab., White Oak, Maryland. (Proj. 675 0502; NOL TR64 125. NOL MDM54. AD–609 970.) OTS prices: HC $2.00/MF $.50.

The NOL Information Retrieval System is a set of programs written in FAP for the 7090. The NOL technical library staff uses this system for document file establishment, maintenance, and searching. A detailed description is given of the methods used to accomplish the above mentioned tasks. (author)

7/65–49 **Information Centers at the Oak Ridge National Laboratory.** November 13, 1964. Francois Kertesz. Oak Ridge National Laboratory, operated by Union Carbide Corporation for the U.S. Atomic Energy Commission, Oak Ridge, Tennessee. (ORNL–TM–996)

The information centers currently operating or about to be established at the Laboratory are reviewed. The factors involved in the creation of the centers and the range of the available services are discussed, mentioning some of the problems thus far encountered in their operation. (author)

7/65–50 **Project Sharp (SHips Analysis and Retrieval Project) Information Storage and Retrieval System: Computer Aspects and Programs.** December 1964. Report Navships 0900–001–4000. Department of the Navy, Bureau of Ships, Washington, D.C. 85 pp.; USGPO price: $.50. The results reported herein were obtained in the course of research supported by the Bureau of Ships, Department of the Navy under Project S–F–007–01–03, Task 0404, at the David Taylor Model Basin, Applied Mathematics Laboratory, Carderock, Md., a field activity of the Bureau of Ships. The report was originally prepared as DTMB Report No. 1923, and edited by Annie E. Cooper and Gilbert R. Gray of the Management Applications Branch, Applied Mathematics Laboratory.

SHARP is a computer oriented information storage and retrieval system developed to resolve some of the problems inherent in the handling, storage and retrieval of scientific and technical literature at the Bureau of Ships Technical Library. The computer is being used to automate to a high degree, (1) bibliographic searches (2) subject matter searches (3) coordinated searches (combinations of both) (4) issuance of library catalog cards and accessions bulletins (5) control of periodicals and journals (6) other aspects under development, such as, complete automatic generic computer searching, automatic posting of descriptive terms by the computer in the indexing procedure, and user interest registers. The system is intended to improve the library's capacity to better serve the user and to function as a management and logistics tool for the library. Details of the computer aspects of the indexing scheme, search strategy, thesaurus, computer programs, present research work are reported. Modifications to the system and future plans are indicated. (author)

7/65–51 **The Automated Approach to Technical Information Retrieval. II.** John J. Nicholas. Naval Engineers Journal, vol. 76, no. 6, December 1964. pp. 923–928.

Describes Project SHARP (*SH*ips *A*nalysis and *R*etrieval *P*roject) being developed by the Bureau of Ships Technical Library. Gives details of the automated periodical inventory control and the present status of the program. (BATR)

7/65–52 **A Seminar on Mechanization in Libraries and Information Centers.** *Technicka Knihovna (Technical Library)*, no. 11, 1964. pp. 350–352. In Czech.

The proceedings of a seminar on mechanization problems in libraries and technical-economic information centers in Czechoslovakia, which was organized by the State Technical Library in Prague in September 1964, are discussed. Reports were given on the mechanization of book circulation, information processing by computers in military intelligence centers, peek-a-boo card systems, KWIC indexing, the mechanized preparation of periodical indexes, large data file organization, and ARITMA data processing equipment. (Doc. Inc.)

7/65–53 **A Time-Tested Functional Filing System (for Pulp and Paper Literature).** January 1965. James d'A. Clark. *Tappi*, vol. 48, no. 1. pp. 57–60A. And *Pulp Paper Mag. Can.*, vol. 66, no. 1. pp. 71–73.

A "vertical" (visible), interest-oriented, physical document-filing system suitable for individuals collecting literature (reports, reprints, data sheets, etc.) relating to the pulp and paper industry is outlined. Legal-size file folders (and cabinets) for various subject categories and subcategories, periodic subdivision of bulky folders, and other practical and common-sense filing procedures are recommended. The particular "classification schedule" for pulp, paper, and board technology developed by the author can easily be changed or modified to suit the personal needs and/or preferences of other workers. (C.L.B.)

7/65–54 **Utilization of Peek-A-Boo Cards.** J. Toman. *Aktualne Problemy Informacji i Dokumentacji (Current Problems of Information and Documentation)*, no. 4, 1964. pp. 11–14. In Polish.

A detailed description of an information storage and retrieval system on peek-a-boo cards, which is in operation in the Building Administration of the Technical and Industrial Construction Bureau in Prague, Czechoslovakia, is given. The file card system was designed to search a collection of documents, mainly journal articles, on microfilms. Each article was indexed, assigned an accession number, and then photographed. Methods of coding information on peek-a-boo cards, the classification system, the data recorded, and the control of information flow are discussed. Advantages of the peek-a-boo card system as compared to a classical card system are: simplicity of structure, low cost, higher speed of retrieval, higher number of access points, and flexibility. (Doc. Inc.)

7/65–55 **The Use of Peek-A-Boo Cards in the Library.** J. Helbich. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei* (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers), no. 2, 1963. pp. 33–40. In Czech.

Experiences in the use of peek-a-boo cards at the Institute for Labor Hygiene and Occupational Diseases in Prague. Czechoslovakia, are reviewed. The cards are used as a store of basic information on literature sources concerned with radiation protection and radiobiology. The card file is to a certain extent a modification of the key-word catalog and supplements the author catalog and the systematic catalog. The use of these cards has been especially efficient in combination with slides and microfilms, since the materials desired can be taken directly from the collection without the necessity of classical cataloging. The alphabetically arranged peek-a-boo card index, based on the principle of subject catalogs, is discussed. (Doc. Inc.)

7/65–56 **Machine Prepared Author and Subject Index. First Report: Method of Index Preparation and Machine Utilization.** F. Luepnitz and M. Luepnitz. *Dokumentation (Documentation)*, no. 4, 1964. pp. 114–116. In German.

Procedures for mechanized preparation of subject and author indexes at the Institute of Agricultural Information and Documentation in East Germany are described. The conventional and mechanized methods of index preparation are compared, and the advantages of the latter are explained. Indexing terms and author names are typed on punch tape from which punch cards are prepared automatically. The punch cards are then sorted on a sorting machine. The next step is a manual operation that includes editing punch cards and classifying them according to a given system. Finally, the cards are tabulated, and the index is printed by offset. The preparation of the indexes by the mechanized method requires 50% less time than does the conventional method, eliminates error sources to a great extent, and reduces considerably the amount of intellectual work required for the preparation of such indexes. (Doc. Inc.)

7/65–57 **Mechanized Preparation of Lists of Periodicals.** J. Svobodova, *Knihovnik (The Librarian)*, no. 8, 1964. pp. 249–251. In Czech.

The mechanized preparation of analytical lists of periodicals at the National Library in Prague, Czechoslovakia, is described. The periodicals are listed by categories, including subject matter, number of issues per year, language, publisher, and other aspects of cultural or economic interest. They are also classified both by the UNESCO Classification System and by the Soviet Classification System for periodicals. The lists are distributed to all Czechoslovakian agencies and institutions that process, publish, or distribute periodicals. The advantages in the mechanization of the above described task are shown. (Doc. Inc.)

Reports of Facilities in Operation—Other Information Centers

7/65–58 **Examples of Punch-Card Applications in the Research Institute of Mineral Oil and Natural Gas in Hungary.** P. Szepesvary. *Tudomanyos es Muszaki Tajekoztatas (Scientific and Technical Information)*, no. 9/10, 1964. pp. 733–747. In Hungarian.

Two examples of storing data on double-row edge-notch cards are described. One example deals with storing chemical and physical data on sulfur compounds. The most important data are punched, while other data are reproduced by drawings (for example, compound structures) or appear in coded form. The identification number of the source from which the data are derived is also recorded. Direct coding is used to denote the classes to which the various sulfur compounds belong. The other example deals with the storage of data by the Laboratory of Mass Spectrography. The laboratory stores the results of its own literature searches, the results of searches conducted by the institute's documentation services, and the results of searches conducted by the special technical branch of the Hungarian Central Technical Library and Documentation Center. The following data are typed on the cards: author, title, source, date of publication, and a short notation of content. The content of the paper is classified and encoded by chemical compounds and processes. (Doc. Inc.)

7/65–59 **The MIT Technical Information Project.** M. M. Kessler. *Physics Today*, vol. 18, March 1965. 28 pp. + 7.

The model of a technical information system involves literature taken from 21 journals in the field of physics. The system, designed and constructed as a prototype operating in a realistic test environment, uses remote consoles having access to a time-sharing computer facility. Programs were developed for a large variety of search and processing techniques in real time as well as for delayed output. The work is supported by the National Science Foundation and in part by Project MAC, the experimental computer facility at MIT which is sponsored by the Advanced Research Project Agency. (BATR)

7/65–60 **Rx for Hospital Records Squeeze.** SSTM, vol. 6, no. 1, January 1965. pp. 24–25.

Describes the Docuform system developed by Documentation, Inc., for Fairfax Hospital, Fairfax, Va.

Reports of Proposed Facilities—Libraries

7/65–61 **Documentation and Dissemination of Technical Knowledge—Means to Achieve The Goal.** A. Erich. *Calauza Bibliotecarului (Librarian's Guide)* no. 5, 1964. pp. 257–259. In Rumanian.

The technical library of the Infratirea plant in Oradea, Rumania, developed a plan for documentation and technical training, which is expected to increase production and lower production costs for 1964. This should be achieved through better information on the production methods used in other parts of the world and by supporting the development of new ideas in technology and labor organization. The information service is handled by a group of engineers and technicians who, among other things, search domestic and foreign magazines. Technical services for workers and joint conferences are also included in the plans. (F.S.)

Reports of Proposed Facilities—Other Information Centers

7/65–62 **Civil Defense Information Systems Analysis (A Feasibility Study of Research Information Exchange).** Volume I of a Two Volume Final Report. January 1965. W. T. Herzog and J. E. Jenkins. Research Triangle Institute, Operations Research and Economics Division, Durham, N.C. Prepared for: Department of the Army, Office of the Secretary of the Army, Office of Civil Defense, Research Directorate (OCD–PS–64–56 OCD Subtask 4631B RTI Number R–OU–158–1).

A feasibility study was performed of a civil defense scientific and technical information system that will insure the ready availability of information to all pertinent OCD elements and to contract research personnel. A brief analysis of scientific information systems is presented as background for a discussion of the present civil defense information system. The potential sources of civil defense information are listed. The civil defense information system is defined to be the relationships or exchange between the sources of civil defense information and the users of this information. Methods for improving this exchange are presented. These suggestions include the establishment of an information analysis center, the printing of a quarterly technical progress review, and the use of standard report format and indexing procedures. To foster standard indexing procedures, a *Thesaurus of Civil Defense Descriptors* is provided. This *Thesaurus* forms the second volume of this two volume report. (authors) (see 7/65–4)

Application of computers to Classification, Indexing, and Text Processing

7/65–63 **A Test of the Factor-Analytically Derived Automate[illegible] Classification Method Applied to Descriptions of Work an[illegible] Search Requests of Nuclear Physicists.** January 1965. Paulin[illegible] Atherton and Harold Borko. Report AIP/DRP–65–1 of th[illegible] American Inst. of Physics. New York, N.Y. and report SP–19[illegible] of the System Development Corp., Santa Monica, Californi[illegible] 11 pp. + appendix.

Using factor-analytic procedures, an attempt was made [illegible] classify nuclear physicists into interest groups and to derive ca[illegible]gories within nuclear physics by an analysis of search reque[illegible] The derived categories were evaluated by subject matter spec[illegible]ists and found to be inadequate. The test helps clarify the reas[illegible] able applications and limitations of automated classification factor analysis. As another exercise, a nuclear physicist logic[illegible] derived a classification of these 77 physicists into thirteen gro[illegible] (authors)

7/65–64 **Automatic Classification for the ASTIA Mathematics Collection.** May 1964. Barry Zimmerman. Master's thesis. Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa. (MSEE–64–22. Contract Nonr55140. AD–610 313.) 48 pp. OTS Prices: HC $2.00/MF $.50.

Information Retrieval of documents based on a series of word descriptors which characterize each document requires the formation of a search description. The user of the library can be greatly aided in the search if he knows which descriptors do not occur together in any document in the file and which descriptors have some relatively high probability of co-occurring within a description. This thesis investigates this absenec and probable presence of co-occurence between pairs of descriptors in the approximately 25,000 document Mathematics section of the ASTIA library. This absence and probable presence is revealed by an Exclusive and an Inclusive Stratification program whose results are presented as a two-level descriptor classification system. The results of this classification are measured by a figure which states the probability that the file contains documents each of which is formed by choosing any two descriptors from on Inclusive Group. Another measure presented is the reduction in the number of possible descriptions of any length. (author)

7/65–65 **Experiments with a New Classification Algorithm.** December 1964. David Senechalle. Report LRC–64–WTM–6 of the Linguistics Research Center. University of Texas, Austin, Texas. NSF grant GN–308.

Experiments leading to the development of a new algorithm for automatic classication are described, and some of the mathematical properties of the problem and the algorithm are investigated. The algorithm locates subsets, termed D-clumps, in an object space consisting of a symmetric connection matrix formed from initial binary object-property incidence data. The D-clumps satisfy criteria of internal cohesion, and of differentiation from their respective complements, that make them useful entities in automatic classification procedures. (author)

7/65–66 **A Method for Producing Journal Indexes (Progress Since AIP/DRP–63–3).** September 1964. Pauline Atherton and Vance Weaver. Report AIP/DRP–64–3 of the Amer. Inst. of Physics, New York, N.Y.

An earlier report, AIP/DRP 63-3 is brought up to date and the progress made since August, 1963 reported. For 11 journals published by the American Institute of Physics and its member societies, this semi-automatic method of producing and printing indexes is employed using typewritten tabulating cards, a xerographic card-duplicator, and a sequential card camera. Costs, advantages and sensitive areas are described. (author)

7/65–67 **Automatic Classification System Users' Manual.** November 1964. Nell Dale. Report LRC–64–TTM–1 of the Linguistics Research Center of the Univ. of Texas, Austin, Texas. NSF grant GN–308.

A FORTRAN IV programming package for experimentation in automatic classification is described from the standpoint of the user. Various options are provided in measuring similarity among individual objects or properties and in locating groups of objects or properties that are similar in some sense (i.e. GR, K or D-clumps). (author)

7/65–68 **Automatic Indexing: A State-of-the-Art Report.** March 30, 1965. Mary Elizabeth Stevens. Monograph 91 of the Institute for Applied Technology, National Bureau of Standards, Washington, D. C. 662 ref. 220 pp; USGPO price $1.50.

A state-of-the-art survey of automatic indexing systems and experiments has been conducted by the Research Information Center and Advisory Service on Information Processing, Information Technology Division, Institute for Applied Technology, National Bureau of Standards. Consideration is first given to indexes compiled by or with the aid of machines, including citation indexes. Automatic derivative indexing is exemplified by key-word-in-context (KWIC) and other word-in-context techniques. Advantages, disadvantages, and possibilities for modification and improvement are discussed. Experiments in automatic assignment indexing are summarized, Related research efforts in such areas as automatic classification and categorization, computer use of thesauri, statistical association techniques, and linguistic data processing are described. A major question is that of evaluation, particularly in view of evidence of human inter-indexer inconsistency. It is concluded that indexes based on words extracted from text are practical for many purposes today, and that automatic assignment indexing and classification experiments show promise for future progress. (author)

7/65–69 **The Use of Context for Correcting Garbled English Text.** Charles M. Vossler and Neil M. Branston. Cornell Aeronautical Laboratory, Inc., Buffalo, New York. *ACM Proc. of the 19th National Conference, Philadelphia, Pa. August 25–27, 1964.* ACM publication P–64. Available from ACM, 211 East 43rd St., New York, N.Y., 10017. Price $5.00.

The paper describes two methods for using context to correct garbled English text. The first makes use of a dictionary of English words containing their probability of occurrence. The second uses letter digram frequencies to roughly approximate English word probabilities. Probabilities of various letter substitutions are obtained from a confusion matrix of the simulated character recognizer whose operation produced the garbling. This information is combined using a maximum likelihood scheme to obtain word recognition or, if only digram information is available, the recognition of word approximations. A bibligraphy and a representative example of the SIR program input and output are contained in an Appendix. (authors)

7/65–70 **SIR: A Statistical Information Retrieval System.** C. D. Parsons. Phillips Petroleum Company, Bartlesville, Oklahoma. *ACM Proc. of the 19th National Conference, Philadelphia, Pa. August 25–27, 1964.* ACM Publication P–64. Available from ACM, 211 East 43rd St., New York, N.Y. 10017. Price $5.00.

This paper describes the techniques and results of an information retrieval system utilizing an IBM 7094 installation at Phillips Petroleum Company. The Statistical Information Retrieval (SIR) system employs a "co-ordinate concept" with a logic based on the statistical probability that a desired document will be abstracted with a relatively high percentage of the identical or synonym key-words used in posing an inquiry concerning the document subjects. This approach to retrieval allows the use of an unlimited vocabulary, thus eliminating the need for a dictionary or thesaurus. The SIR programming incorporates a unique computing technique, vector manipulations and a search strategy which permit the system to operate efficiently on a large-scale computer. In order to test the methods empirically, computer programs were written, and experiments were run using textual material from various sources. Besides a rather limited comparison of the "dictionary" and "digram" methods on material from a children's primer, a test was made of a combined system on material from newspaper articles and from a book on psychology. (author)

7/65–71 **SDI–5, An Advanced System for Selective Dissemination of Information.** A. B. Barnes, A. A. Briggs, J. Gauss and A. Resnick. IBM—Advanced Systems Development Division, Yorktown Heights, N.Y. *ACM Proc. of the 19th National Conference, Philadelphia, Pa. August 25–27, 1964.* ACM Publication P–64. Available from ACM, 211 East 43rd St., New York, N.Y. 10017. Price $5.00.

Continued experimentation and operational experience have been incorporated into the programming of a new IBM 1401–7090 Data Processing System for Selective Dissemination of Information (SDI) now serving over 1200 scientific, technical and administrative personnel. The most significant new features are:

1. Machine printing of all bibliographic information including abstracts and matching index terms.
2. Control by the user of the type and quantity of information he receives.
3. Segmentation of user's interest profiles into distinct interest areas.

4. Weighting of user index terms.
5. Extensive analysis routines to monitor user and system performance.
6. Several procedures for automatically updating user interest pofiles.
7. Machine-controlled vocabulary guide for both user and document profiles. (author)

7/65–72 **Computer To Microfilm, "A New Era."** 1964. Harry G. Cooper. NMA Proc. (see 7/65–1).

Economics and convenience of microfilm recording directly from computers for business and scientific applications. Paper will cover actual systems in use today and cost comparison. (author)

7/65–73 **High Speed Reproduction of Library Cards Through Microreproduction Techniques.** 1964. Allen B. Vearner. NMA Proc. (see 7/65–1).

The production of library catalogue cards has been a bottleneck in the work flow patterns of libraries. Printing methods have been inefficient, wasteful, or unsuited to the technical procedures of some libraries. Microreproduction techniques has been developed which can satisfy the requirements of high quality, rapid delivery and minimal waste. This highly automated method of printing more than 5,000 cards per hour, is currently being used to generate over a million cards each year for a large research library. Production is expandable five-fold with no increase in equipment. (author)

7/65–74 **Computer Input from Printing Control Tapes.** 1964. Lee Ohringer. *TAGA Proc.* Tech. Assoc. of the Graphic Arts, Rochester, N.Y. pp. 304–316.

Project UPGRADE (supported by DOD and NSF grants) at the Univ. of Pittsburgh Computation & Data Processing Center is progressing along three lines: (1) development of a system for recording currently published technical literature in computer-compatible form, using paper tapes punched during linotype, monotype, and/or phototype printing; (2) development of a computerized typesetting system for book production; and (3) development of *PENELOPE* (Pitt Natural Language Process), a computer language for handling nature language text instead of numeric data to enable efficient programs to be written by people with minimum programming training. (C.L.B.)

7/65–75 **Probabilistic Pairs and Groups of Words in a Text.** April-June 1964. A. R. Meetham. LANS, vol. 7, Pt. 2. pp. 98–106.

The report is a continuation of "Preliminary Studies for Machine Generated Index Vocabularies," A. R. Meetham (1963) *Language and Speech,* 6, 22. It assumes that documents employ words from particular groups connected with their subject matter, and discusses four methods, two of them new, for finding the groups. They are picked out from a word list by using a word-word binary matrix to represent the associations between pairs of words. In an evaluation, the method which consumes least computer time turns out also to be the best. (author)

7/65–76 **A Factor Analytically Derived Classification System for Psychological Reports.** Harold Borko. System Development Corporation, Santa Monica, California. PMOS, no. 20, 1965. pp. 393–406.

A sample of approximately 1,000 abstracts was obtained from *Psychological Abstracts* and keypunched for computer processing. Based upon a frequency distribution of words in these abstracts, 150 tag-terms were selected. These terms were intercorrelated on the basis of their co-occurrence in documents, and the resulting matrix was factor analyzed. The factors were interpreted as representing classification categories. These were compared with, and shown to be similar to, the APA classification system. The study demonstrates that it is possible to determine the basic dimensions of a collection of documents by an analysis of the words used in their abstracts. (author)

7/65–77 **Secondary Key Retrieval Using an IBM 7090–1301 System.** D. R. Davis and A. D. Lin. IBM Corporation,San Jose, California. CACM, vol. 8, no. 4, April 1965. pp. 243–246.

The secondary key retrieval method involves the preparation of secondary storage lists from primary data records. Search requests are satisfied by logical operations on appropriate lists, producing a complete set of addresses of primary records relevant to the request. Experimental results are presented and a comparative analysis is given. (authors)

7/65–78 **The Future Role of Computer-Based Systems in the Graphic Arts Industry.** 1964. Morton Morgenstern. *TAGA Proc.* Tech. Assoc. of the Graphic Arts, Rochester, N.Y., pp. 297–303.

Computer capabilities for potential applications in the printing-publishing industry may include: capture of information during keyboarding, optical scanning of manuscripts for translation into computer language, display terminals for direct proofreading and entry of corrections or alterations, driving of photocomposing machines to produce page negatives, direct transmission of copy to an electrostatic press, machine translation of books into several languages, information retrieval for writers and reporters having high-speed access to archival data, publication of custom-tailored monographs based on automatic retrieval techniques, and supervisory-management functions, such as cost estimating, inventory control, and production scheduling. (C.L.B.)

7/65–79 **Computer Applications for Commercial Printing Reviewed.** January 1965. Theodore C. Collins and Harold F. Drury. *U.S. Dept. of Commerce, BDSA Quart. Ind. Dept.* pp. 3–7.

Advantages of computers as management and production tools in commercial printing and publishing are reviwed, and computer installations in newspapers as of Nov. 1964 are listed. (C.L.B.)

Subject and Linguistic Analysis—General

7/65–80 **Computational Linguistics: Procedures and Problems.** January 1965. W. P. Lehmann. Report LRC–65–WA–1 of the Linguistics Research Center, University of Texas, Austin, Texas. NSF grant GN–308.

The necessity for a more complete understanding of language as the basis for machine translation and computational linguistics is stressed. Other benefits which will result from this longterm research—including information retrieval and automatic classification —are also mentioned. (author)

7/65–81 **Basic Format Offers Uniformity in Cataloging Library, Management and Engineering Data.** May 4, 1964. Burton E. Lamkin and Allan D. Pratt. IBM General Products Division, San Jose, Calif. (TR 02.304.) 11 pp., 1 fig., 4 refs.

The use of computers and unit record equipment for processing of information *about* information continues to increase. Typical examples of such processing are information retrieval systems, KWIC (keyword in context) indexes, procedure manual indexes, SDI (selective dissemination of information), library cataloging processes, and parts catalogs. One characteristic which is common to all these systems is that of recording, in machine-readable form, "document derivative data for each document (broadly defined to include books, periodicals, reports, etc.) that is placed in the system. This report represents a basic format for uniform cataloging of a wide variety of information for both library and nonlibrary applications. (authors)

7/65–82 **History and Use of the KWIC Index Concept.** 1964. Marguerite Francis Fischer. (Thesis.) San Jose State College, San Jose, Calif.

The KWIC Index is defined and described. Its history and that of its literature are given, as well as a brief biography of its originator, Hans Peter Luhn. Variants of the KWIC Index, such as the Bell Telephone Index, the KWOC, and the WADEX, are discussed, with examples given. The KWIC Index is evaluated in

reference to Taube's criteria: extrinsic and intrinsic. It is likewise evaluated as a current awareness tool (with the accompaniment of SDI to enlarge its usefulness) and as a retrospective searching tool (with editing to make it acceptable).

Suggestions are made for its improvement. better titles, the device of part titles, manipulated titles, use of a thesaurus, accompanying classification information, author index interspersed with title index, cross-references, improvement of type-face, code changes, a single lokup index, and understanding by users of its nature. The KWIC Index is then viewed in relationship to a total information system and is deemed worthy of having its place in an eventual national index. (author)

7/65–83 **Classification of Information. Part I: Principles of a Controlled Vocabulary.** J. Toman. *Metodika a Technika Informaci (Methodology and Technique of Information)*, no. 9, 1964. pp. 1–10. In Czech.

The larger the store of documents or information, the more elaborate and precise are the requirements of the system and the file organization. Knowledge of the principles of classification is one of the basic prerogatives for the effective organization of an information storage system. An analysis of two basic approaches, the method of subject headings and the method of systematic classification, reveals the existence of two basic principles of data organization. Problems of a spontaneously developed and a controlled vocabulary are described along the Uniterm indexing methods and thesauruses. (author)

7/65–84 **Automatic Translation of Languages.** Silvio Ceccato. ISSR, vol. 2, no. 3, December 1964. pp. 105–158.

Describes the work of the Centro di Cibernetica e di Attivita Linguistiche, University of Milan, Italy, in the areas of language analysis and mechanical translation from Italian into English. (BATR)

7/65–85 **Communications, Semantics, and Information Systems.** L. S. Hill. *Journal of Industrial Engineering*, vol. 16, no. 2, March-April 1965. pp. 131–135.

Semantic considerations militate against the successful accumulation, storage, and retrieval of information for management control systems. Some case examples of semantic interference are given PERT cost as a frame of reference. (BATR)

7/65–86 **Expanding the Editing Function In Language Data Processing.** L. B. Doyle. System Development Corporation, Santa Monica, California. CACM, vol. 8, no. 4, April 1965. pp. 238–243.

In automatic abstracting, citation indexing, mechanical translation and other such procedures, editing is required whenever the automatic method leaves something to be desired. This paper discusses the economy of editing as a function of the amount of condensation of text in language processing operations, and then contends that editing can be regarded as an opportunity rather than as an unwelcome necessity.

Heavy editing, which goes beyond mere correction and improvement of computer output, is exemplified by the use of a concordance in preparing a survey article or lecture. Other opportunities for heavy editing are described, chief among them being interpretation and expansion of computer output in such processes as factor analysis. Applications are described, such as the quick, unbiased evaluation of a large volume of incoming mail or telegrams, yielding summary reports not possible for either humans or computers to produce along. (author)

7/65–87 **Mathematical Linguistics.** F. Kiefer. *Orszagos Muszaki Konyvtar es Dokumentacios Kozpont, Idoszeru Muszaki Dokumentacios Kerdesek (Hungarian Central Technical Library and Documentation Center, Current Problems of Technical Documentation)*, no. 3/4, 1963. pp. 143–172. In Hungarian.

Mathematical linguistics is based on the similarity between grammatical and mathematical structures. Different mathematical methods are applied to description and analysis. According to the main mathematical disciplines there have been developed set-theoretical, statistical, mathematical-logical, information-theoretical, and algebraic models of languages. In a set-theoretical model developed by O. S. Kulagina, the word, the environment, and the grammatically correct sentence are basic notions; language is defined in three parameters: (1) set of words; (2) set of environments; and (3) set of grammatically correct sentences. Kulagina syntactic units of sentences. The set-theoretical definition makes also developed the theory of configurations corresponding to the it possible to trace complicated sentences back to simpler ones. The article also describes and analyzes the mathematical-logical model of Y. Bar-Hillel and J. Lambet, the algebraic model, which utilizes the formal properties of differently defined languages, the theories of Chomsky and his school, the methods of statistical linguistics developed by Herdan, Zipf, and Mandelbrot, and the methods based on teh information theory. (C.B.L.)

Specific Classification and Indexing Systems

7/65–88 **Introduction to the Classification of Metallurgical Literature.** Dr. A. Uhlman. *Tudomanyos es Muszaki Tejekoztatas* (Scientific and Technical Information), vol. 3, April, 1964. pp. 196–209. In Hungarian.

Problems and procedures of the classification of metallurgy in the UDC system are discussed. Charts of classes and subclasses of metallurgical terms are available. The importance of alphabetical indexes is growing. In these indexes the entries are followed by the decimal numbers which correspond to the ones in the chart. As comprehensive charts are not adequate anymore, special charts are needed and these items should also be included in the alphabetical index. Linguistic problems arise in translations, not only because dictionaries—even the special ones—are incomplete, but also because the logic of different languages is not the same. German, English, and Hungarian metallurgic terms are compared in great detail. The method of assigning decimal numbers to complex terms is described such as 669.721.5'74'5 for Mg, Mn, Z alloys. (C.B.L.)

7/65–89 **Mechanized Search of Technical Documents.** D. L. Kazachkov. *Mekhanizatsiia i Avtomatizatsiia Proizvodstva*, vol. 19, no. 1, January 1965. pp. 39–41. In Russian.

Descriptors used for the classification of machine construction parts and coding of aperture cards are described. Mekhanizatsiya poiskov tekhnicheskoy dokumentatsii. (BATR)

Coding and Notation

7/65–91 **Indexing Newspapers and Periodicals.** I. Onodi. *Tudomanyos es Muszaki Tajekoztates (Scientific and Technical Information)*, vol. 1, January 1964. pp. 9–14. In Hungarian.

On January 1, 1964, a uniform method of identifying newspapers and periodicals was introduced at the recommendation of the United Postal and Telecommunication Organization of the Socialist Countries by the member nations of that organization. Groups of five-digit codes in the number range 10,000 to 99,999 have been assigned to the member countries (e.g. Hungary is assigned 25,000–29,999). Each country allots one of its assigned numbers to each of its publications. This system is intended to facilitate uniform cataloging, simplify administrative procedures, promote mechanization, and enable standardization of coding systems. The paper analyzes further applications of this system in library operations. (author)

7/65–92 **A Cluster of Algorithms Relating the Nomenclature of Organic Compounds to Their Structure Matrices and Ciphers.** G. M. Dyson. ISSR, vol. 2, no. 3, December 1964. pp. 159–199.

General methods were devised for the computer conversion of the names of organic compounds to structural matrices from

which the notation and molecular formulas can be automatically deduced. The procedures involve the systematic use, by a cluster of algorithms, of the stored data concerning the nonsystematic morphemes and semantemes of organic nonenclature. (BATR)

7/65–93 **Computer Program for the LINCO System.** H. Bouman. JCHD, vol. 5, no. 1, February 1965. pp. 14–24.

The previous name LOCUS has been changed to LINCO. An outline is given for a computer program for transforming the notation for a chemical compound, made by clerical staff directly from the structural formula and containing arbitrarily allotted numbers, into a unique, unambiguous notation, which can be translated into other codes. (BATR)

7/65–94 **A Graph-Theoretic Algorithm for Matching Chemical Structures.** Edward H. Sussenguth, Jr. JCHD, vol. 5, no. 1, February 1965. pp. 36–43.

Gives a generalization and comprehensive description of the algorithm, proofs of convergence and realted topics, and applications other than chemical retrieval systems. (BATR)

7/65–95 **A Chemical Structure Storage and Search System Developed at Du Pont.** D. J. Gluck. JCHD, vol. 5, no. 1, February 1965. pp. 43–51.

Gives the details of the topological system developed at Du Pont, including input computer programs, each algorithm, and economics of the system. (BATR)

Special Processing Languages

7/65–96 **Optimizing Retrieval Results with Man-Machine Interaction.** February 1965. Robert M. Curtice and Victor Rosenberg. Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa. NSF grant GE-2569. 24 pp.

This paper discusses the development of a retrieval search strategy based on two fundamental principles: man-machine interaction and the generation of index terms based on their co-occurrence as they are assigned to documents. The search request is reformulated by the user with a machine generated list of index terms related to his request. The procedure consists of a series of computer programs providing the user with various alternative methods of formulating the question and the search. The reformulated request is designed to increase the recall and specificity of the retrieval. (authors)

7/65–97 **Automated Information Systems in Planning, Control and Command.** Andrew Vazsonyi. *Management Science*, vol. 11, no. 4, February 1965. pp. B2–B41.

Discusses the problems of automation, on-line, real-time information systems, control of airline reservations, man-machine communication consoles, PERT, and automated teaching of languages. (BATR)

7/65–98 **A Computer User-Oriented System.** George D. Montillon. *ACM Communications*, vol. 8, February 1965. pp. 117–124.

A computer language system was developed which makes possible fast preparation of management reports. The system requires initial preparation of large data banks containing data in elementary form. Use of two special languages, EXTRACT and MATRAN, permits selective extraction of any data dubset, efficient processing through any computational sequence, and flexible presentation of results in either tabular or graphical form.

Equipment and Material Descriptions

7/65–99 **Progress Through EDMS.** 1964. Otis F. Wooley. NMA Proc. (see 7/65–1).

During the past five years, EDMS progress has resulted in the release of Specifications and Standards covering 35mm microfilming, 105mm filming, tabulating and aperture card preparation. The experience gained through the use of these documents has dictated that new and broader guidelines are required to support new programs. Currently broader requirements are being developed to permit a higher degree of mechanization in the Engineering Data Micro-Reproduction Systems area. (author)

7/65–100 **Scratch Resistant Parameters of Microfilm Emulsions.** 1964. William S. Suydam. NMA Proc. (see 7/65–1).

The scratch resistance of silver microllm emulsions has been of some concern for all users. A study of processing methods as trey affect the scratch resistance (hardness) of the film emulsion sheds some light upon the problems. The new ASA test "Methods for Determining the Scratch Resistance of Processed Photographic Film" was used as a guide for comparing various films and processing baths combinations. Some current beliefs have been found to be contradictory to facts brought out by this study. (author)

7/65–101 **Electrostatic Layout Drafting Table.** 1964. W. D. Pegram. NMA Proc. (see 7/65–1).

The equipment can best be described as a Xerographic powder image transfer and fixing apparatus that, while not limited thereto, is adapted to add discrete bits of information to drawings or documents. The equipment was developed to, among other things, improve the microfilmable quality of notes, tables and other data normally applied to drawings by typewriter or pressure sensitive tapes. (author)

7/65–102 **Microdensitometry in a Microfilm Program.** 1964. H. C. Frey & H. E. Rubin NMA Proc. (see 7/65–1).

7/65–103 **Unitized Film Processing.** 1964. Arthur W. Kutchera. NMA Proc. (see 7/65–1).

Microfilm can now be rapidly processed in unitized form to a high quality, permanent micro-image. There are many variables which must be considered to create such a product. The technique of processing and the control of specific parameters are essential to the rapid access unitized film form. (author)

7/65–104 **World-Wide Exchange of Scholarly and Industrial Microfilm Programs.** 1964. John Hensel. NMA Proc. (see 7/65–1).

An aid in scientific work, the microfiche has been rediscovered in our country, one of many contributions from across the ocean. Continued investigation into better ways of industrial efficiency through microfilm and its results have found wide acceptance in other lands. What can be done to accelerate the active use of microfilm among us?. Industry must give consideration to international norms and market conditions. (author)

7/65–105 **A High Speed Planetary Camera.** 1964. John Devitt, and Harry Bennett. NMA Proc. (see 7/65–1).

This paper describes an automatic microfilming camera using a specially designed conveyor and planetary camera to give high speed operation and high resolution. The camera was especially developed for the microfilming of freight waybills for the conversion to invoices and reproduction. It is used, however, for the high speed accurate microfilming of various types of documents. Vacuum holddown, automatic light control and automatic camera masking are employed. A single operator can produce up to 50 exposures per minute. (authors)

7/65–106 **The New Publishing Prodigy—Microfiche.** 1964. A. L. Baptie. NMA Proc. (see 7/65–1).

The phenomenal growth of microfiche use in this country has occurred within a short time. What will be the long-range impact on the microfiche industry and on publishing? Presentation, including slides, explores: (1) the concept of the unit microform; (2) use for industrial catalogs; (3) use for technical reports; (4) activities of federal government information agencies; (5) NMA microfiche standards; (6) microfiche equipment. (author)

7/65–107 **The Levitation Principle—Its Application and Progress.** 1964. Gunter Schmidt. NMA Proc. (see 7/65–1).

A very short historical background on immersion and spray processing is given leading to the levitation principle introduced in 1961, and its application in Levitron and Super Levitron process machines. (author)

7/65–108 **The Relationship Between Microfilm Negatives and Hard Copy Print Out.** 1964. Terry Wilson. NMA Proc. (see 7/65–1).

Ultra-violet projection of microfilm makes possible a microreproduction system which will compete with other processes for reproducing engineering drawings. The work engendered the study of the relationship between master negatives and the final diazo printout and an overall study of the relationship between each generation of a microreproduction system, with a view to losing no information in the final printout. Methods of achieving this and a suggested specification are indicated, which falls within the existing D.O.D. specification. (author)

7/65–109 **Microreproduction Requirements for Information Processing.** 1964. J. E. Crow. NMA Proc. (see 7/65–1).

The proven value of technical information retrieval systems has resulted in a phenominal increase in both the number and size of document collections by both large and small organizations. Microreproductions make an important contribution to solution of the storage, retrieval and communications problems within this field. However, their use in this vast market, although considerable, has fallen short of the possible potential, resulting in a loss to both the microreproduction and information groups. This paper discusses basic reasons for this situation and offers potential solutions. (author)

7/65–110 **Microscopic Spots: A Progress Report.** 1964. R. W. Henn and D. G. Wiest. NMA Proc. (see 7/65–1).

The recently reported microscopic spot formations found in rolls of processed negative microfilm occur most commonly as circular spots in the background and leader. Occasionally, some degradation of image lines occurs. The silver is missing from the attacked area and has been redeposited in the surrounds. Little damage to information recorded on microfilm has occurred. However, the possible occurrence of these spots has important implications in the archival storage of microfilm. The spots are described and illustrated and techniques for inspection of microfilm collections are given. Procedures for improved storage and a gold protective processing treatment are described. (author)

7/65–111 **Rapid Processing of Recordak Micro-File Film at Elevated Temperatures.** 1964. J. C. Barnes. NMA (see 7/65–1).

7/65–112 **Basic Guidelines for Setting Up an Engineering Drawing Microfilm Program.** 1964. Wn. P. Burleigh, Jr. NMA Proc. (see 7/65–1).

Initial investigation during which the economy and practical feasibility is tested, affected groups are informed of plans, samples are made for evaluation, and other important steps taken. Systems Installation including establishing proper standards, assigning responsibilities, testing the system and other major aspects. Follow-up to insure smooth operation and ultimate expansion. The paper will conclude with a discussion of some dozen common pitfalls. (author)

7/65–113 **Vesicular Systems for High Resolution Generation Printing.** 1964. M. G. Anderson. NMA Proc. (see 7/65–1).

Various Kalvar film types are evaluated interms of the carbon step tablet method of determining printing density. Combinations are demonstrated for static 3rd, 4th and 5th generation printing with silver halide and diazo films as first and second generations. Vesicle size in the application of the high Callier Q factor of Kalvar vesicular system, test methods, ultra-high projection resolutions and microfilm generation printing as in Mil. Std 9868 are discussed. Generation prints obtained on a roll-to-roll printer-processor are presented. (author)

7/65–114 **The Successful Marriage of 35mm and 105mm Film Systems.** 1964. Robert K. Wendel. NMA Proc. (see 7/65–1).

Companies often submit 35mm microfilm to Department of Defense agencies and attempt to use existing engineering drawings. The existing engineering material is often rejected before it gets to the microfilm camera for quality and consistency of lines, positioning of formats, etc. Should the drawing be completely redrawn? Is there a method that can accomplish the task faster and far more economically? The 105mm miniaturization system has as one of its design purposes the photographic reproduction and restoration of engineering drawings. This support function by 105mm film has taken many directions and the following examples are indicative of the savings possible in both time and money in the preparation of existing drawings for the 35mm microfilming operation. (author)

7/65–115 **Quality Control At The Library of Congress.** 1964. Donald C. Holmes. NMA Pro. (see 7/65–1).

The Library of Congress is filming some 100,000 deteriorating bound newspaper volumes as a preservation and space-saving measure, and some 700 different current newspaper titles in lieu also being microfilmed. Because the originals in most cases are of binding; deteriorating reference books and other materials are discarded under these programs, it is imperative that the microfilm produced be of the highest quality, and the Photoduplication Service of the Library has developed an extensive and intensive program of quality control to meet the need. (author)

7/65–116 **Microfiche Specification Review For Auto-Viewing And Printing.** 1964. John Rutkus. NMA Proc. (see 7/65–1).

As the Government Facilities Dissemination Programs generate increased quantities of microfiche, the need for Enlarger Printers and Viewers to automatically create hard copy and facilitate auto-viewing becomes more critical. The present N.M.A. Microfiche Specification has inherent limitations that will be discussed as regards to its utilization for automated Viewers and Enlarger Printers. (author)

7/65–117 **The NASA Microfiche Program.** 1964. Melvin S. Day. NMA Proc. (see 7/65–1).

The NASA Microfiche program has been developed from the concept of assuring the local availability of complete documents to scientific and technical users at the time such users learn of the existence of the basic information through the NASA abstract journal. The Microfiches are diazo negatives, lending themselves to single image reproduction, or entire microform reproduction. of a U.S. federal agency specification for Microfiche which addi-. Cooperation with other Government agencies has led to adoption tional agencies are now using in their own expanded technical information programs. (author)

7/65–118 **Microfilm in the Graphic Arts.** 1964. T. G. Maynard. NMA Proc. (see 7/65–1).

Use of the principle of electronic character generation for high speed composition at rates of 17,000 characters per second of copy for printing of publications. This includes the general approach of computer programming to produce magnetic tape to drive a character generator with CRT output producing positive microfilm without hard copy. The automated system employed projects the film to paper negatives for offset plate making and printing on large presses for production of catalog publications. (author)

7/65–119 **Protection and Rejuvenation of Microfilm Rolls and Aperture Cards.** 1964. Lawrence L. Steele. NMA Proc. (see 7/65–1).

World War II and the growth of the Electronic and Missile Age resulted in millions of drawings requiring frequent retrieval. However, scratches and abrasions in aperture cards and roll film caused confusion by competing with draftsmen's line and important numerals were obliterated. Large microfilm users found a protective method and a rejuvenation system used by the motion picture industry the answer to their problem. Slides and details of exhaustive tests will be included together with illustrations of a new microfilm protective treatment machine. (author)

7/65–120 **Electronic Data Retrieval and Microfilm.** 1964. Ernest P. Taubes. NMA Proc. (see 7/65–1).

Information can be stored in large quantities electronically and retrieved in microseconds, but the conversion to permanent electric typewriters unless the characters are displayed on oscillo-humanly readable characters is limited by the capabilities of scope screens and microfilmed. The various media presently available to display alphanumerical information including the possibility of using computer information to make actual designs of standard component parts assembled for special applications such as road construction, bridge, etc. and the subsequent microfilm of this information as a permanent record. (author)

7/65–121 **What Job Cost Control Can Do for Microfilming Operations.** 1964. Jos. T. Simone. NMA Proc. (see 7/65–1).

A proper operating cost system can provide powerful management information and a tool to insure profitable business operations. Analysis of the following questions: (1) How much value is there in knowing weeks or months ahead of time that your department or business is going to show an operating loss or that you are not going to operate on objective? (2) Why can such information be useful on a daily basis in order that you can take action immediately? (3) How can you fully evaluate the cost of setting up your own microfilming and reproducing facilities as opposed to using the service company? (author)

7/65–122 **Educating Top Management to The Importance of Microfilm as a Business Systems Tool.** 1964. Rodd Exelbert, Joseph L. Kish, and Peggy Sholtz. NMA Proc. (see 7/65–1).

7/65–123 **Modern Microfilm Production-Through the Eyes of a Production Manager.** 1964. Robert H. Holliday. NMA Proc. (see 7/65–1).

A brief history of production methods and some salty definitions of sales, developments and research, quality control and production as seen from the eyes of each. Analysis of a good production setup brings some good pointers on a number of items such as production control methods, job descriptions, training methods, product flow and work analysis, communications, application of new equipment and techniques, cost accounting and production, with appropriate examples. Finally, what does the future hold in the way of new equipment, techniques, supplies and methods to get there. (author)

7/65–124 **Imagination, Ingenuity—Significant Elements in Effective Microfilm Program.** 1964. Margaret M. Weis. NMA Proc. (see 7/65–1).

Useful tricks, interesting techniques, essential standards and routines developed during 22-year microfilming program covering indexing, handling, control and destruction of almost 200 varieties of business records and related microfilm copies—in small or large quantities—on 100-foot rolls of 16mm film at 24:1 reduction—for protection of Vital records and/or retention of long-term records on microfilm with destruction of originals. (author)

7/65–1 **Data Management System Requirements.** 1964. Parker H. Daggett, Jr. NMA Proc. (see 7/65–1).

7/65–126 **Microscopic Spots: A Processing Effect.** R. W. Henn, D. G. Wiest and B. D. Mack NAMN, no. 73, December 1964. pp. 92–96.

7/65–127 **Microfiche, Microfilm, and Hard Copy—Problems and Prospects for the Research Worker.** William R. Hawken. NAMN, no. 73, December 1964. pp. 97–107.

7/65–128 **Revised Microcopy Resolution Test Charts.** NAMN, no. 72. October 1964. pp. 85–88.

NBS recently revised the microcopy resolution test charts it produces for use in measuring the resolving power of microfilm systems. The new charts, which go on sale Dec. 1, 1964, provide for a wider range of spatial frequencies than were available on the old charts. (author)

7/65–129 **Information Processing and Mechanization in Technical Libraries and Information Centers.** Compiled by Anna Vejsova, Prague, Statni, Technicka Knihovna (State Technical Library), 1963. 131 pp. Volume 2 of *Exchange and Experiences*, Reviewed by J. Podzimek in *Technicka Knihovna (Technical Library)*, No. 3, 1964. pp. 86–87. In Czech.

This is a summary of material presented at two seminars held by the State Technical Library in Prague. The first part is an informative presentation of available organizational tools and mechanical equipment for information processing. It contains brief descriptions of punch card equipment and of automatic computer systems. The second part describes new developments in Czechoslovakia and other eastern block countries: experiences with alphanumerical ARITMA machines in Germany, catalogue processing by the photoholerite method in Hungary, use of alphabetical tabulators ARITMA 320 and other ARITMA machinery in Czechoslovakia, etc. The third part deals with computer systems and their applications. A survey of reproduction techniques and equipment is also included. (M.H.G.)

7/65–130 **Equipment for Viewing Peek-A-Boo Cards.** J. Havrankova. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers)*, no. 2, 1963. pp. 41–42. In Czech.

Equipment for viewing peek-a-boo cards is described. It is an illuminating device similar to that used in making photocopies. The device is recommended for facilitating the finding of documents. (Doc. Inc.)

7/65–131 **Information Retrieval for Paper Industry Management.** Peter W. Sherwood. *Pulp Paper Mag. Can.*, vol. 66, no. 1, January 1965. pp. 82–83.

In a brief, but salient review of manual and mechanized IR systems (edged-notched and Peek-a-boo cards, punched-card sorters, FLIP, Filesearch, Ferranti-Packard Rapid-Access Lookup, Filmsort, Minicard, Magnavue, Verac, Walnut, Microcite), the use of computers is predicted to increase, but technical and economic obstacles are pointed out. Computer expenditures for IR will rise from $20 million in 1963 to $90 million in 1968. The SRI study is quoted, showing that annual operating costs for computerized search systems start to pay at 40,000 items/month input and 250 searches/month, or 15,000/month input and 1000 searches/month. The bottleneck of human document selection, analysis, and indexing may be solved with the advent of optical scanners and page-to-tape converters. (C.L.B.)

7/65–132 **Progress Report, Research on Aging Blemishes on Microfilm.** C. S. McCamy. NAMN, no. 74, February 1965. pp. 133–7.

7/65–133 **Microfische and Research Reports.** A. L. Baptie. NAMN, no. 74, February 1965. pp. 138–44.

7/65–134 **Possible Applications of the Punch Card System in Processing Documents on Metallurgy.** D. Simandl. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers)*, no. 2, 1963. pp. 57–58. In Czech.

Potential applications of punch cards for processing literature on metallurgy, in general, and the use of the handsorter analyzer in providing a small special supplementary information store in such areas as various methods for refining different types of ores and steel alloys are described. The codes that were designed supplement the Universal Decimal Classification System and provide for a quicker access to the specialized information store. In addition, the use of keypunch machines is discussed and numerous experiences in their use are described. (Doc. Inc.)

7/65–135 **"Automata Konyvtaros"** (The Automatic Librarian), *Nepszabadsag (Peoples' Freedom),* February 1964, p. 8. In Hungarian.

A Hungarian invention, the Automatic Librarian, makes it possible for one man to handle an entire library. It is based on a punch card system that performs 40 different services. One of the services performed is the marking on the lending slip the date of lending and expiration. During the current year there will be several such machines on trial. Next year, wholesale production is foreseen. (D.R.M.)

7/65–136 **Xerography—Polish Device Pylorys KS2.** M. Shour. *Mechanizace a Automatizace Administrativy (Mechanization and Automation in Administration),* no. 10, 1964. pp. 246–247. In Czech.

The Xerox method of document reproduction is explained and the Polish machine (Pylorys KS2) is described. This machine is capable of producing copies of the original size, enlargements of 1:2, and reductions of 2:1. The processing time is three minutes. Copies can be made in three colors. The maximum number of copies from one plate is six. Power requirements are 3,000 watts; the weight is approximately 330 pounds. Improvements of the Xerox equipment and processes by the manufacturer of these devices in Lodz, Poland, are reported. (Doc. Inc.)

7/65–137 **Punch Cards in Mechanized Information Processing.** J. Bukovsky. *Metodicke Letaky (Methodological Pamphlet),* no. 44, 1964. pp. 1–13. In Czech.

The use of punch cards for mechanized information processing by computers in general and by ARITMA machines in particular is discussed. Various types of codes, punching methods and procedures, and the use of punch tapes are discribed. The functions of data processing centers, potential uses of punch card systems in library operations, and keypunching and processing costs are analyzed. (Doc. Inc.)

7/65–138 **Microfilm Techniques in Information and Documentation Centers.** F. Kretschmer. Spirit, J., and Thannabauer, V., *Metodika a Technika Informaci (Methodology and Technique of Information),* no. 1, 1964. pp. 30–48. In Czech.

The authors report in detail their experiences with process cameras of both the stationary and portable type. The optimization of exposure is investigated. Different types of enlarging and developing equipment are discussed. Photographic material such as films, hard copy paper, developers, fixers, and weakening and intensifying solutions for copying is classified and analyzed. The article also describes a method for determining sensitivity of films, procedures for photographing, developing, and retouching negatives and positives, and procedures for preparing slides. (author)

7/65–139 **Mechanization of Book Circulation.** O. A. Krivtsun. *Bibliotekar (The Librarian),* no. 12, 1964. pp. 49–52. In Bulgarian.

Various equipment and machines used in western countries for mechanizing book circulation in libraries are described. These include machines for checking out books by transferring information from reader cards to circulation cards, in particular those produced by Dikman, Gaylord, and Bro-Dart-Sysdak. Another large category of machines registers book loans on control lists. There are various types of such machines: photographic, electromechanical, audio, machines that print information on special cards, and punch-card processing machines. Research results on the effectiveness of such machines are reported. (Doc. Inc.)

7/65–140 **Means for Mechanization and Automation of Work in Libraries.** M. Chytil and J. Nesicky. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei. (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers),* no. 2, 1963. pp. 6–30. In Czech.

A survey is presented of available means for information processing. The peek-a-boo card system and edge notch cards and their analyzar are discussed. The alphanumerical Aritma punch card machines (keypuncher, verifier, sorter, alphanumerical tabulator, and duplicator) are described. Their applicability for data processing in libraries and scientific-technical information centers is investigated and cost estimates are given. The fundamentals of computer components, such as the memory unit, the arithmetic unit, the central processing controller, and input and output devices, are discussed. In addition, a survey of computers in use in Czechoslovakia and associated cost estimates for information processing are given. Practical examples of the application of peek-a-boo and edge notch cards and machines in information processing are presented. (Doc. Inc.)

Use and User Studies

7/65–141 **Technology Transfer and the Flow of Technical Information in a Large Industrial Corporation.** March 1965. Curtis P. McLaughlin, Richard S. Rosenbloom, and Francis W. Wolek. Grad. School of Business Administration, Harvard University, Boston, Mass. NASA grant NsG–253–62. 37 pp. Appendices in a separate vol.

This is a user study of the acquisition of useful technical information by scientists and engineers in five divisions of a large industrial corporation. More than 1,200 instances of the acquisition of such information were described on self-administered questionnaires by 430 respondents. Additional data were gathered by personal interviews with more than 30 respondents. The survey shows that the respondent's intention in searching for information, the nature of the source which proved useful, and the medium which joined source and recipient are closely related aspects of the communications process. The report discusses the pattern of relationships among these characteristics and considers how it is influenced by personal factors, such as experience, and situational factors, such as the type of work done in the engineering section. Explanations for the observed relationships are suggested on the basis of interview data concerning organization structure and policy, geography, and the character of the relevant technologies. Thus, the transfer of technical information is depicted as a complex social process, one in which technology, personal characteristics, organizational values and structure, and the beliefs of significant professional subcultures interact to influence the patterns of behavior. (author)

7/65–142 **Theoretical and Methodological Considerations in Undertaking Innovations in Scientific Information Exchange.** January 1965. Report 12 of the Amer. Psychological Assoc. Project on Scientific Information Exchange in Psychology. Washington, D.C. NSF grant GN–281.

Attempts to solve informational problems can have far-reaching effects not only upon scientific information exchange but also upon the ultimate form of the discipline and the conduct of scientific work. While it is premature to lay out an ideal system of dissemination and to consider all innovations in terms of their moving the system towards the desired state, it is still necessary that careful consideration be given to possible effects prior to undertaking changes in means of disseminating, storing and retrieving scientific information. (authors)

7/65–143 **Do Present Information Services Serve the Engineer?** John Sayer. DAPR, February 1965. pp. 24–25, 64–65.

7/65–144 **Putting Information to Work.** T. J. Devlin. CEPR, vol. 61, no. 3, March 1965. pp. 35–37.

Esso R & E's information services provide working engineers with abstract journals plus documents as requested; 1,200 professional employees make 85,000 requests per year. (author)

7/65–145 **Information Storage/ Retrieval: Is it Working?** Saul Herner and W. F. Johanningsmeier. CEPR, vol. 61, no. 3, March 1965. pp. 23–29.

Survey shows about one-third of readers use keyword index or abstracts in A.I. Ch. E. publications; most use them for current awareness. The therarus is used to a smaller extent; need to publicize two tools cited. (authors)

7/65–146 **How Engineers Use Literature.** R. A. Davis. CEPR, vol. 61, no. 3, March 1965. pp. 30–34.

They put much of what is available to good use, but need to know more about what's on hand; manufacturer's catalogs rank quite high in usage. (author)

7/65–147 **Technical Libraries: Users and Their Demands.** 1964. Margaret Slater. Aslib, London SW 1, England. (Z675.T3 SL15t Over.) 126 pp.

This study presents the results of a pilot investigation of the use made of technical libraries and information units in industrial units in industrial firms, government laboratories, academic institutions, and learned societies. The three main aims are determination of user demands, classification of user groups and their demands, and estimation of the significance of the above for librarians. (BATR)

7/65–148 **Open Stacks and Catalogs.** J. Lobotkova. *Ctenar (The Reader),* Prague, no. 6, 1964. pp. 221–223. In Czech.

Statistics show that an open stacks system stimulates the interest of readers and increases the circulation of books. The author discusses the merits of the open stacks system and disproves the allegations that it could lead to undisciplined habits or to undue influence of the library staff on the readers' selections. The open stacks must contain the basic and essential works of all fields, including belles-lettres. The books should be arranged according to an uncomplicated classification system. Since the open stacks do not include all the books of the library, they must be supplemented by systematic author and subject catalogs. (Doc. Inc.)

7/65–149 **A Survey of Services Offered Authors By Biological Editors.** Domenic A. Fuccillo, Jr. Oak Ridge National Laboratory, Oak Ridge, Tennessee. Reprinted from *BioScience* vol. 14, no. 2, 1964. pp. 23–24.

Since authors often need information on services provided by editors to prepare manuscripts effectively, the Conference of Biological Editors through its Committee on Graduate Training queried its membership to obtain pertinent data. This report presents results of a survey on the scope and kind of services they offer authors. The study was designed, however, to gather information that might be useful to two groups: (1) graduate students in biology, and other new authors submitting manuscripts to journals; and (2) editors interested in services offered by their colleagues. More than 100 CBE members were asked 15 questions. A summary and discussion of the replies of 87 respondents is presented here. (author)

7/65–150 **Mooers' Law (Some Remarks on the Psychology and Sociology of Information Utilization.** P. Gyore. *Tudomanyos es Muszaki Tajekoztatas (Scientific and Technical Information),* no. 9/10, 1964. pp. 748–763.

According to Mooers, no information retrieval system will be used until the user finds the state of being informed more discomforting than the state of not being informed. There are deep psychological reasons behind this law. Most people prefer to maintain the status quo, but at the same time they wish something new would happen. New information often excites and exerts a disturbing influence. Creative people are not frightened by this feeling, but those performing usual routine activities are often deeply affected. Methods are described for supplying information to the latter group. Creative ability and the capacity for assimilating information are related to each other, thus the development of creative ability will improve the receptivity to accepting information. The ecology of information is discussed. Among others, it studies the relationship among certain categories of suppliers, and the relationship among these categories and the various systems of information dissemination. (Doc. Inc.)

7/65–151 **The Information Needs of Scientific Research.** A. Falvay. *Tudomanyos es Muszaki Tajekoztatas (Scientific and Technical Information),* no. 7, 1964. pp. 541–549. In Hungarian.

The need for information in scientific research is analyzed and evaluated. Each research activity has individual characteristics and the requirements for information are varied. Various types and sources of information pertinent to the specific requirements of scientific research are investigated. All phases of the research activities should be coordinated with information and documentation agencies. Results of the efforts of these agencies do not always correspond to the actual requirements. Suggestions are made on the methods of improving operations and achieving better coordination. The organization of a special institution to deal with these problems is proposed. (Doc. Inc.)

Design, Testing and Evaluation of Information Systems

7/65–152 **Evaluation of Science Communication Systems.** February 1965. Lawrence S. Papier. Report CRDLSP–4–65 of the Directorate of Technical Services, U.S. Army Edgewood Arsenal Chemical Research and Development Labs., Edgewood Arsenal, Md. 49 pp.

This report presents seven areas of investigation into which evaluators must probe to determine the overall worth of an information retrieval system (IRS). These are system objectives, user requirements, external relations, system properties, system efficiency, technical aspects, and operational aspects. Technical and operational personnel should participate in the evaluation. A determination of user requirements includes the following: (1) the types of communication necessary to fulfill user requirements; (2) user requirements in terms of system-response properties; (3) the ability of the system to provide the proper communications to the user; and (4) a quantitative and intuitive evaluaion of fulfillment of user requirements. Two resultant listings (properties of system response and types of communications) form the basis of a coordinate grid that may be expanded to any degree necessary. (author)

7/65–153 **Recommended Approaches to Design of the U.S. Army Engineering Data and Information System (EDIS).** December 1964. Stanley A. Goldberg. Army Research Office, Office of the Chief, Research and Development, Washington, D.C. Rept. no. EDIS 2. (AD–453 737L.) 26 pp.

All release of this document is controlled. All certified requesters shall obtain release approval from Director of Army Technical Information. Office of the Chief of Research and Development, Army Research Office, Wash. 25, D. C.

7/65–154 **An Analysis of the Multiple Instantaneous Response File.** December 1964. Ronald J. Ferris. Rome Air Development Center, Griffiss AFB, N.Y. (RADC–TDR64 457. Proj. 4594. AD–610 131.) 27 pp.; OTS Prices: HC $2.00/MF $.50.

The functional capability of a device called the Multiple Instantaneous Response File (MIRF) is clarified. The MIRF is a document surrogate, storage and retrieval device which employs a mechanically alterable, read only content addressable, parallel search memory. Organization of the memory permits the user of the file to be on-line in addition to providing automatic search refinement functions. Some attention is given to the evaluation of MIRF, but this involves essentially a review of the

evaluation problem. This report does not discuss the hardware aspects of the MIRF but it does briefly mention the applicability of memories having the same basic features. (author)

7/65–155 **Evaluation of the Impact of a Citation Index in Physics.** September 1, 1964. Ben-Ami Lipetz. Report AIP/DRP Cl-3 (1964) of the American Institute of Physics, New York, N.Y. 64 pp. + App.

An experiment was conducted for the purpose of evaluating the impact of a citation index to physics literature upon the literature-use habits of physicists. A citation index was specially prepared in which it is possible to look up a known reference in either *The Physical Review* or the *Journal of Applied Physics* to learn whether, and precisely where, it was cited in four of the leading Soviet physics journals which are published in English translation by the American Institute of Physics. It was hypothesized that the index would help United States physicists to learn of Soviet articles with subject contents similar to the known articles they would start with, and would therefore tend to increase their utilization of the physics translation journals.

The impact of the experimental citation index was measured principally by means of statistics on utilization of Soviet physics translation journals at over 100 libraries throughout the United States. The statistics were in several categories, reflecting borrowing activity, photocopying activity, and browsing activity. They related to each of the four journals used as sources in the citation index, and to an additional four Soviet physics translation journals which were excluded from the index to serve as controls. Statistics were supplied for a period of months preceding and following the controlled distribution of the experimental citation index. The index was distributed without advance notice to all subscribers to *The Physical Review* and *Journal of Applied Physics* in a restricted, self-contained geographic region which accounts for five percent of the domestic subscriptions to American Institute of Physics publications. Library statistics from this test region were compared with statistics from the remainder of the country in order to evaluate the impact of the citation index.

Results obtained from the various measurement and evaluation techniques used in the experiment are in good agreement. They indicate that the citation index produced a definite, but quite small, increase in source literature utilization by the physicists who received the index. Only a small fraction of the physicists who received the index actually tried to apply it to practical literature search problems; those who did so apply it were almost unanimously enthusiastic about the usefulness of the citation index approach.

It is concluded that the citation index technique *in its present state of development* can be applied with highly beneficial results to some of the literature search problems of physicists. The relatively poor over-all acceptance and utilization of the experimental citation index by physicists indicate that this type of reference aid will be evolutionary, not revolutionary, in its contribution to the improvement of literature retrieval in physics. (author)

7/65–156 **Study to Determine the Applicability of the Solomon Computer to Command and Control. Vol. I. Information Storage, Retrieval and Communication System Control.** Final rept. October 1964. Franklyn M. Rybak. Westinghouse Electric Corp., Baltimore, Md. (Contract AF19 628 2846. Proj. 2801. ESD TDR64 184 Vol. 1. AD–454 765.) 194 pp.

During an 8 month's study, research was performed to determine the applicability of the SOLOMON I computer, a highly parallel, network organized machine, to processing functions encountered in command and control systems. Specific tasks pursued were: (1) investigation into the benefits to be gained in information storage and retrieval systems through file maintenance and search with SOLOMON I; (2) consideration of the control of large communication networks through real-time simulation on SOLOMON I; and (3) study of the control and processing functions in radar sensor subsystems as well as in radar oriented weapons control systems. To illustrate the potential of parallel organized computers, especially SOLOMON I, in command and control systems, specific application examples are provided. It is concluded that computers employing the organizational concepts of SOLOMON I can yield speed advantages over conventional sequential machines, of up to several hundred times. In addition, application of redundancy to the basic SOLOMON I system can yield substantial reliability and cost advantages over conventional duplex configurations for critical command and control functions. (author)

7/65–157 **An Evaluation of the Interlibrary Loan Service of Wayne State University Medical Library: 1963-1964.** September 1964. Gwendolyn S. Cruzat. Report 5 of Wayne State University School of Medicine Library and Biomedical Information Service Center, Detroit, Mich. 26 pp.

7/65–158 **A Technique for Utilizing the IBM or the RCA Random-Access Mass-Memory Devices to Store the Data Base of a Command and Control Information Processing System.** November 1964. Paul E. Friedenberg and Thomas S. Walton. David Taylor Model Basin, Washington, D.C. (DTMB–1917. Proj. SS192 001, Task 7160. AD–610 211.) 13 pp.; OTS Prices: HC $1.00/MF $.50.

Two different mass memories are studied for their suitability to store the data base in a command and control information-processing system. The IBM unit employs metal disk packs whereas the RCA unit uses magazines of plastic cards for the recording medium. However, analogies in their logical capabilities make it possible to use identical methods of file organization. It is shown that the equipment can be effectively utilized to process formated files on a random-access basis. Compact indexes are generated so that items can be quickly located either by name or by content. The times required for updating and searching such files are analyzed, and the advantages to be gained over conventional techniques are indicated. (author)

7/65–159 **Impact of Microfilm on Retrieval Hardware.** 1964. Alan C. Root. NMA Proc. (see 7/65–1).

When documents could be read by the human eye, human hands could file and retrieve them. Aperture cards and microfiche users will need equipment for access to the files. Retrieval hardware is being developed to save the user access time; it will be a point-of-use machine and will allow infiling at random within 100 cards. Future retriever modules will run card into reader-printer, automatically re-file it, as the file remains within the system. (author)

7/65–160 **New Uses for Microfilm in Industry.** 1964. E. C. Carlson. NMA Proc. (see 7/65–1).

Many new time and money saving microfilm applications are being put to use by industry today. This presentation reviews in detail the Receiving-Inspection, Assembly Line and the Reproduction Reference File applications. Also discussed will be a variety of other applications, such as, personnel records, Industrial X-Rays and Engineering Support Documents. (author)

7/65–161 **Tentative Bibliography on Evaluation of Information Systems.** February 18, 1965. Compiled by Madeline M. Henderson. National Bureau of Standards, Research Information Center and Advisory Service on Information Processing, Washington, D. C.

7/65–162 **Summary of Study Conference on Evaluation of Document Searching Systems and Procedures.** October 2–3, 1964, Marvel Hall, American Chemical Society Building. National Science Foundation, Washington, D.C., 10 February 1965.

Copies of the summary can be obtained from: Clearinghouse for Federal Scientific and Technical Information (402), Institute for Applied Technology, National Bureau of Standards, 5285 Port Royal Road, Springfield, Virginia. 20235. Price: $2.00 for a hard copy, $.50 for a microfiche.

This small conference was arranged by the National Science Foundation (NSF) to review work on the testing and evaluation of document searching systems and procedures and to consider

promising directions for future work in this area. The scope of the conference was limited to systems and procedures for searching for documents that satisfy particular search specifications. There was no intent to limit the notion of system—examples might be the tools and procedures used by reference libraries and information centers to locate wanted documents; indexes and other tools and their modes of use by users of various kinds; the staff, equipment and procedures that comprise partially mechanized document searching systems; and more fully automatic document-analysis-and-searching procedures under development.

7/65–163 **Key Words In—and Out of—Context.** B. B. Lane. SPLB, vol. 55, no. 1, January 1964. pp. 45–46.

Ten periodical indexes were examined to determine how suitable KWIC indexing might be for areas of knowledge other than chemistry. Results indicate that in science and engineering the titles of articles usually describe or imply the contents of the articles. In non-technical fields, titles reveal the contents less frequently; in a general index, they are indicative less than half the time.

7/65–164 **The Measurement and Evaluation of Reference Service.** Samuel Rothstein. LIBT, vol. 12, January 1964. pp. 456–472.

Reviews the trends in the measurement and evaluation of reference services. Author warns that reference librarians, in failing to provide the means for accurate judgment on their contribution in library service, run the risk of having their work undervalued or ignored.

7/65–165 **Some Comments on Information Retrieval—Personal Observations.** Gary R. Martins. DTMN vol. 10, no. 12. December 1964. pp. 24–27.

The single, complex purpose of mechanized IR is to make available *unusually attractive* library services. Seven *unusual services,* viz., convenient access, rapid results, *relatively recent* data availability, suggestive indexing and reference, broad and/or deep stockpiles of information, auto-abstracting, and *creative* retrieval, are discussed, indicating modest results achieved so far, formidable obstacles remaining, and some promising research approaches. Cost/effectiveness ratios govern many of the sub-problem areas. There is need for tests to evaluate the worth of indexing and request-processing schemes, and also for criteria of quality in library accessions. Perhaps the best approach to heuristic, docile sub-systems for useful, economical IR is further research in "on-line" data processing (man-machine interaction). (C.L.B.)

7/65–166 **Problems of Information Searching.** G. Hovanyi. *Tudomanyos es Muszaki Tajekoztates (Scientific and Technical Information),* no. 9/10, 1964. pp. 764–781. In Hungarian.

There are three characteristics which differentiate information searches from those of bibliographic and literary searches. These are *deeper, broader,* and *more methodical systematic searching.* An example of the systematic method of searching is the related series of operations that can be performed on search terms and notions. These include the replacement of subject and aspect terms by descriptions, their categorization, the use of roles and links, concept analyses, coordination, and storage in the search system files. Examples are given of these operations in various existing search systems. In addition, a table is presented illustrating which of these systems, in general, are adaptable to executing these various operations. (Doc. Inc.)

7/65–167 **How to Organize Information Systems.** John Dearden. *Harvard Business Review,* vol. 43, no. 2, March-April 1965. pp. 65–73.

This article suggests an approach to systems organization which will help business to take advantage of future developments in modern data-processing equipment and techniques. (BATR)

7/65–168 **The Automated Approach to Technical Information Retrieval.** John J. Nicolaus. NVEJ. Pt. I. October 1964. pp. 715–727. Pt. II. December 1964. pp. 923–928. (See 4/65–76.)

7/65–169 **A Dokumentacio Gazdasagossaga** (Cost Analysis of Documentation). B. Csendes. Publ. Orszagos Muszaki Konyvtar es Dokumentacios Kozpont—Modszertaoni Kiadvanysorozata *(Hungarian Central Technical Library and Documentation Center—Methodological Publication Series),* Budapest, Hungary, no. 1, 1964. p. 211. In Hungarian.

The expenditures for and the effectiveness of documentation work are analyzed. The purpose of this analysis is to survey the status and expected development of operational methods in the field of documentation in foreign countries, and to propose standards and criteria for planning the time and cost requirements of documentation work in Hungary. The role of documentation within the framework of research and production is reviewed. Estimates are made of expenditures for production, with respect to research and development, as well as of the relationship between these expenditures and expenditures for documentation. Detailed time and cost analyses for documentation operations are presented. The documentation center, its work, and organization are described, and a cost budget for an average size facility is drawn up. It is emphasized that documentation is not only a service but also a part of the production process and must, therefore, follow the rules of production mechanics. There must be a reasonable relationship between the services of and the requirements for documentation centers, and a strict line of demarcation must be drawn between documentation operations and those which are not. This book includes a bibliography with 120 entries. (Doc. Inc.)

7/65–171 **A Comparison of Documentation Systems.** P. Vasarhelyi. *Tudomanyos es Muszaki Tajekoztatas (Scientific and Technical Information),* no. 9/10, 1964. pp. 782–789. In Hungarian.

The work of Western Reserve University in comparing various documentation services under strictly defined conditions, as observed during a student tour, is discribed. Test methods, developed by Cleverdon for the Cranfield Project, are discussed. The possibilities of mathematical modeling and analyzing various documentation and indexing systems are investigated. (Doc. Inc.)

7/65–172 **Stipulations for Improving the Effectiveness of the Work with Technical Literature.** I. Diesenberger. *Technicka Knihovna (Technical Library),* no. 12, 1964. pp. 353–362. In Czech.

Reasons for the ineffectiveness of the work with technical literature, evidenced by the discrepancy between the amount of technical information utilized and the amount of technical information searched and processed, are investigated. Various possibilities for increasing effectiveness and efficiency of information processing and utilization are analyzed. A more careful study of users' requirements is suggested that would place emphasis on more recent information material, the specialization of services, the standardization of procedures, and an active participation of information users in systems design. Methods for controlling the effect of information services on the nation's economy and the development of science and technology are proposed. (Doc. Inc.)

7/65–173 **Analysis of the Effectiveness of Work with Technical Literature.** I. Wiesenberger. *Technika Knihovna (Technical Library),* no. 11, 1964. pp. 329–340. In Czech.

The author investigates the effectiveness of the results produced by libraries and information centers in their work with technical literature and compares various methods of work with respect to their efficiency. Productivity, effectiveness, and profitability of the human effort in this field are defined. A measure of the effectiveness of information prepared on request is the time it saves administrative and research workers—the requestors—by relieving them of the necessity of conducting information searches. Various types of directed information services, such as technical documentation, literature searches, and state-of-the-art studies are listed and their effectiveness is estimated in the percentage of time-saving to the users. The effectiveness of general information dissemination services is measured by how the ideas contained in the information are transformed into practical applications by the workers, technicians, scientists, and other users

of such information. An analysis of this type of information service includes special publications, technical films, and special exhibitions. A quality evaluation point system is proposed. According to this evaluation, special films and special lectures have the highest effectiveness ratings. Limitations of such evaluation systems are recognized, and suggestions for quality improvement of information services are made. (Doc. Inc.)

7/65–174 **Analysis of Work As a Source of Technical Improvement in Library Operations.** T. Krzyzhewski. *Bibliotekarz (The Librarian),* no. 10, 1964. pp. 303–307. In Polish.

The need for technical improvement in library operations in Poland is explained. Very little progress was made in this respect during the last twenty years. The work of technical libraries, in particular, is investigated, and suggestions are made for simplifying procedures for the inventorizing and accounting of acquisitions as well as for the cataloging of books. The number of entries on catalog cards should be reduced because a considerable amount of the recorded information now required is superfluous. Eventually, reduced and reproduced title pages could be pasted on cards and used for cataloging. It would be highly desirable to establish a Central Technical Library in Poland and to produce union catalogs of technical literature. The introduction of an open-shelf system would also simplify the librarians' work. Broader use should be made of Kardex files and other modern office equipment. (Doc. Inc.)

7/65–175 **Documentation and the Efficiency of Library Operations.** C. Dima-Dragan. *Calauza Bibliotecarului (Librarian's Guide)*, no. 4, 1964. pp. 225–228. In Rumanian.

This is an analytical study of the efficiency of library operations and the role of documentation in library work. The study covered the libraries of the province of Oltenia. It was concluded that: (a) libraries should make better use of audio-visual devices such as charts, posters, slides, etc.; (b) departments of library operations should become research laboratories of the library science; (c) the efficiency and the practical value of the methods of library operations should manifest itself in a steady increase of library activities. The importance of on-the-spot guidance to librarians is emphasized. Documentation needs for library operations are described. (F.S.)

7/65–176 **The Derivation of Statistical Mean Values on the Amount of Labor Required for Various Types of Activities in the Scientific Libraries of East Germany.** W. Dux and G. Ith. *Zentralblatt fuer Bibliothekswesen (Central Journal of Library Science)*, no. 9, 1964. pp. 528–539. In German.

Time studies were made in the German Library in Leipzig (Deutsche Buecherei Leipzig) and twelve large university libraries of the following library activities: entering titles in the alphabetic catalog, recording new acquisitions, recording incoming periodicals, and ordering catalog cards. The purpose of these studies was to investigate the possibility of establishing standards for comparing the efficiency of various libraries and measuring labor productivity. Data analysis proved that the chosen approach to the solution of the problem was correct and that more extensive efforts in this direction would be justified. (Doc. Inc.)

7/65–177 **The Possibilities of the Mechanization of Technical Information and the Use of Punch Cards.** P. Vasarhelyi. *Tudomanyos es Muszaki Tajekoztatas (Scientific and Technical Information)*, no. 3, 1964. pp. 177–184. In Hungarian.

A centralized system is proposed for producing punched cards for company libraries. Disadvantages of independent punching by each library are explained, and the advantages and methods of extending mechanical operations into all phases of library processes are shown. Analysis is made of the results of time studies and of coding and classification problems. (author)

7/65–178 **Scientific Libraries and the Increase of Work Productivity.** W. Kaltschmidt. *Zentralblatt fuer Bibliothekswesen (Central Journal of Library Science)*, no. 9, 1964. pp. 515–525. In German.

Scientific libraries need to increase productivity since demands on their personnel, resources and services continue to increase. The effectiveness of scientific libraries has a strong impact on industrial productivity, research, and development. The author discusses the possibilities of standardization, mechanization and rationalization of various library operations. Mechanization of processes requires review and modification of existing procedures. Studies of interlibrary loans disclosed an urgent need to reduce the time required to process an order. The author concludes that it is necessary to devise a method to measure library effectiveness, which is a difficult task, since it is not indicated simply by the number of books loaned. Rather, the productivity index should be a composite measure of quantity, quality, and costs of the varied tasks the library performs, as well as the facilities and services made available to the users. (Doc. Inc.)

7/65–179 **Library Statistics.** M. Bilek. *Bibliotekarstvo (Library Science),* no. 3, 1964. pp. 21–33. In Serbo-Croatian.

The lack of adequate library statistics in Yugoslavia is discussed. After mentioning examples of unreliable statistical reports sometimes published, the regular statistical returns of different types of libraries are described. These include: the numerical relationship between holdings and the population, the number of books used in relation to the population, the number of books used by various age and occupational groups, and the number of professional books used. Scientific libraries prepare data and statistical tables that show the distribution and use of books in the various scientific fields. The National Library and the university libraries have special regulations for statistical reporting. Rules for preparing library statistics in the following areas are presented: special library materials, special groups of users, professional and age distribution of readers, yearly book acquisitions, equipment and facilities, and financial management and economy. Adequate statistics must be furnished before it will be possible to analyze future needs and plan the future programs of Yugoslavian libraries. (Doc. Inc.)

7/65–180 **Methodik Des Recherchierens und Recherchemittel. *(Methods and Tools of Information Search).*** G. Schmoll. VEB Verlag fuer Buch- und Bibliothekswesen, Leipzig, 1964, p. 65. Series of the Library of a Documentalist, vol. 5. In German.

Information search is defined as an activity whose purpose is to locate the sources of information, to identify the information pertinent to the request of the user or to uncover specific facts and/or data. The result of an information search might be a special bibliography or a state-of-the-art study. The author investigates the importance and the role of such activities in the documentation and information field, information sources to be taken in consideration, methods of approach to information search, qualifications of the information specialist required for a successful search, and the tools available to him. A list of the most important reference works, bibliographies, directories, libraries, and information centers is concluded. (Doc. Inc.)

7/65–181 **Problems in Using Punch Card Machines.** E. Mater. *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers),* no. 2, 1963. pp. 60–65. In Czech.

Problems that present difficulties in the development of mechanized information processing are discussed. Emphasis is placed on the problem of foreign languages and the transcription of foreign language words relative to the existing capacity of the machines for such a reproduction. Various Aritma punch card machines and their components are discussed in the light of their applicability in libraries and in technical and economic information centers. (Doc. Inc.)

7/65–182 **A Standard Measure of Library Productivity.** J. Kovacsics and K. Dux-Nagy. *Magyar Konyvszemle (Hungarian Book Review),* no. 3, 1964. pp. 225–238. In Hungarian.

The authors criticize the traditional methods of library statistics which compare the libraries either on the basis of absolute

figures or percentages which are not related to the library size. A new method, intensity indicators, is proposed. Intensity indicators are dimensionless numbers expressing the ratio of circulation, acquisition, and other figures of a given library to the corresponding figures of a hypothetical average library. The figures of the hypothetical average library are derived as statistical means of all the libraries in the country. Intensity indicators are additive and their sum gives a measure of certain aspects of library work for the entire country. (Doc. Inc.)

Training in Documentation

7/65–183 **Study of Information Science Curricula: 1. Library Schools.** Albert A. Melkonian and Joseph C. Donohue. *Parameters of Information Science.* pp. 57–59. (see 4/65–1)

Developing curricula in information sciences will be influenced by needs of students, as well as by special intellectual interests of faculty. To help identify subject matter and emphases common to all, and to offer guidance to potential students, a study was made of existing and projected programs in ALA-accredited graduate library schools. Plans are outlined for continuing study of information science curricula developments in library schools, as well as other institutions. (authors)

7/65–184 **Blueprints for Accredition.** E. J. Humeston, Jr. and and John F. Harvey. *Parameters of Information Science.* pp. 61–65. (see 4/65–1)

Continuing their discussion (ADI Annual Meeting, 1963, Short Papers, Part 2) of ADI accreditation of instructional programs in information science, the authors indicate the National Commission on Accrediting's criteria for recognition of accrediting agencies and present in some detail a number of suggestions for the development of complete educational standards for information science programs, touching upon faculty, admission and graduation requirements, budget, quarters and equipment, and other phases of the program relative to accreditation. (authors)

7/65–185 **Information Science Instruction in ALA Accredited Library Schools.** Santina Maria Isabella. *Parameters of Information Science.* pp. 67–71. (see 4/65–1)

This paper, based on the data collected for the author's MBS thesis, is an up-to-date summary of the present and proposed information science content of the ALA accredited library schools' curricula. Each library school is individually discussed. The study represents the most complete study in this area of information science education. (author)

7/65–186 **Curriculum Development in Documentation and the Information Sciences.** Robert S. Taylor. *Parameters of Information Science.* pp. 31–37. (see 4/65–1)

7/65–187 **Manpower in Science Information.** William D. Hitt, Robert S. Kohn, and Roger C. Van Buskirk. *Parameters of Information Science.* pp. 39–41. (see 4/65–1)

A report on the first phase of a two-phase program designed to provide a base of information to aid the National Science Foundation make sound decisions with respect to science-information manpower questions. Describes the methodology used and the questionnaires developed for the prospective survey, and presents tentative conclusions. (authors)

7/65–188 **An Undergraduate Course in the Information Sciences.** Robert S. Taylor. *Parameters of Information Science.* pp. 73–77. (see 4/65–1)

In the Fall Semester (1963–64), Lehigh University offered an undergraduate course as an introduction to the information sciences. The course had two objectives: to interest students in the subject; and to impose some organizing structure on a subject in flux. The course was made up of six topics, in addition to an introduction: 1. historical aspects; 2. subject analysis; 3. experimental methods; 4. language data processing; 5. behavioral aspects; and 6. systems concepts. The course is discussed in terms of its objectives, with an assessment of future action. (author)

7/65–189 **Review and Critique of Undergraduate Course in the Information Sciences.** Report No. 1 on the Curriculum for the Information Sciences. March 16, 1964. Robert S. Taylor. Center for the Information Sciences. Lehigh Univ., Bethlehem, Pa. 15 pp.

7/65–190 **Training of Documentalists.** Report of the meeting of FID/TD. Warsaw, May 21–23, 1964. Int'l. Fed. for Documentation. The Hague, the Netherlands. August 1964. 46 pp.

7/65–191 **International Exchange of Information Specialists.** J. Toman. *Metodika a Technika Informaci (Methodology and Technique of Information)*, no. 10, 1964. pp. 37–39. In Czech.

The international exchange of information specialists among member nations of the International Federation for Documentation, as part of the training program for documentalists, is surveyed. The importance of such an exchange to professional advancement is explained. Various possibilities for financing the exchange programs are investigated. Since some countries have problems in allocating foreign currency funds, international organizations, such as UNESCO, should contribute more generously to the exchange programs. Member nations of the IFD should inform the secretariat as to the number of exchange specialists they would be willing to accept and where they would like to send theirs. A training program for such exchange specialists should be worked out. Suggestions for drafting such a training program are made. (Doc. Inc.)

7/65–192 **Felsofoku Dokumentacios Tanfolyamok a Szovjetunioban *(Advanced Documentation Courses in the Soviet Union)*,** Publ. by the State Technical Library and Documentation Center, Budapest, 1964. 53 pp. In Hungarian.

This book contains the detailed program and lecture outlines of advanced documentation courses in the USRR. The program was prepared by a task force of the Standing Committee on Documentation of the National Committee on Technical Development. It contains outlines and bibliographies of the following courses: the theory and organization of scientific and technical information; the mechanization and automation of data processing, including electronic data-retrieving machines; the economics of scientific and technical information services; bibliographic work and familiarization with information sources; editing of technical publications; the Universal Decimal Classification System and its application in scientific and technical information; and new techniques in reproduction. (L.D.)

7/65–193 **Training of Documentalists in Scandinavia.** *Technicka Knihovna (Technical Library)*, no. 11. 1964. p. 352. In Czech.

Various training programs and courses for documentation specialists in Sweden, Denmark, Norway, and Finland are outlined. A historical survey is given of the development of these activities. The organizations responsible for training programs are listed. The topics of lectures given during the courses are described. These include documentation techniques, classification systems, information processing methods, literature searches, and other topics. The Scandinavian Council for Applied Research has been organizing international seminars and courses since 1958. (Doc. Inc.)

Documentation—General

7/65–194 **Mathematical Theories of Relevance with Respect to the Problems of Indexing. Report No. 1: The Formal Basis of Relevance Judgements.** July 9, 1963. Donald J. Hillman. Center for the Information Sciences, Lehigh University, Bethlehem, Pa. NSF grant GN–177. 21 pp.

The foundations of a mathematical theory of relevance are described, and a connection established with the topological theory

of retrieval constructed in Report No. 6 of *Study of Theories and Models of Information Storage and Retrieval.* A basic matching predicate is given a concrete interpretation *via* an index term association matrix developed by Giuliano and Jones, and the mathematical properties of matching are set forth. On the basis of the interpreted predicate, index terms are sorted into groups of conceptually related items, concepts themselves are defined, and actual techniques for assessing relevance are described.

7/65–195 **The Information Explosion and the Science of Language.** 1963. J. W. Perry. Arizona University, Tucson, Arizona. Grant AF AFOSR62 289 AFOSR 64 0561 (AD–434 419). 13 pp.

Reprint from The Proceedings of the 6th Annual Institute in Technical and Industrial Communications, Colorado State Univ., Fort Collins, July 1963, pp. 106–118. (Copies supplied by DDC.)

7/65–196 **Acsi-matic Technical Manual OB Collation System. tem. Vol. 5.4. Retrieval Subsystem.** August 1964. 1 vol. Radio Corp. of America, Bethesda, Md. Contract DA49 0830SA2338 (AD–447–185).

This report is written for the system analyst and describes the ways in which information is retrieved from the various ACSI-MATIC tape and disc files. It includes a discussion of the basic design features of each retrieval technique and a description of how each type of retrieval is performed. Unless otherwise stated, all descriptions in this report refer to the system capability as of March 6, 1964. (author)

7/65–197 **Computer Programming Techniques for Intelligence Analyst Application.** Quarterly rept. no. 1. August 1964. Thomas J. Watson Research Center, Yorktown Heights, N. Y. Contract AF30 602 3303, Proj. 4594 RADC TDR64 233 (AD–605 267). 48 pp.; OTS prices: HC $2.00/MF $.50.

Contents: statistical prediction, discrimination, and classification technique; integrated computer-oriented information retrieval techniques; multiprogramming techniques for intelligence information processing; computer console input and display (textual and graphic); automated program debugging techniques.

7/65–198 **Information Retrieval Thesauri.** Eugene Wall. November 1962. Engineers Joint Council. New York, N. Y.

Purposes of an information retrieval thesaurus are enumerated in Part I and the characteristics required to achieve the purposes are described. The information retrieval thesaurus is shown to combine the techniques of library cross-referencing and thesaurus term-grouping. A brief history of the information retrieval thesaurus is provided, and reasons for calling such a tool *thesaurus* are set forth. In Part II, a detailed example is presented wherein results obtained by searching with and without a thesaurus are quantitatively compared. (author)

7/65–199 **Information Sciences 1963.** Annual rept. January 1964. Air Force Office of Scientific Research, Washington, D. C. AFOSR 64 0101. (AD–600 110). 47 pp.; OTS price $4.60.

7/65–200 **Information Science Seminars.** Final rept. December 18, 1963. Lehigh University, Bethlehem, Pa. Grant Nonr G00029 63, Task NR049 185 (AD–425 464). 5 pp.; OTS price $1.10.

7/65–201 **Probability and the Library Problem.** January 1962. M. E. Maron. RAND Corp., Santa Monica, Calif. P–2471 (AD–606 197). 22 pp.; OTS prices: HC $1.00/MF $.50.

7/65–202 **Applications of the Galois Connection to Information Systems.** September 15, 1964. R. F. Barnes and R. W. Robinson. Report 9078–1 of the Itek Corporation, Lexington, Mass. Contract AF 49(638)1229.

Our inquiry into the quantitative aspects of the concepts of similarity as applied to a simple variety of library systems has led to the consideration of a large family of association schemes. Each scheme is uniquely determined by a particular transformation function—a map from the set of similarity coefficients for terms (or documents) to the set of similarity coefficients for documents (or terms). The transformations considered preserve the property of nonnegativity for sets of similarity coefficients, which allows us to show that there is a set of coefficients left fixed by any chosen transformation. This set of coefficients depends on the chosen transformation and represents exactly the degree of association demanded by consistency with the chosen transformation. Thus, for any transformation, we arrive at a set of similarity coefficients which corresponds to it.

An iterative method was advanced for the computation of our fixed set of similarity coefficients, and the particular analytic questions remaining center upon this method. It has not been shown that the iteration always converges, though experience indicates that convergence is the rule. If the iteration converges at all, it converges to a fixed point as desired. Experience further indicates that the fixed point obtained is independent of the starting set of coefficients, providing these coefficients are nonnegative, except possibly on a set of measure zero.

Further work should center upon the unproven conjectures about the iteration method, and upon the appropriate choice of transformation for particular applications. The transformation which should be used in the context of a library retrieval scheme, for instance, would depend on such various library parameters as the total number of documents in the collection and the average number of descriptors per document. Also, the purpose of the search is likely to affect the choice of transformation. For example, a search can be made more or less strongly associative by varying μ in the range $0 < \mu < 2$. Thus, the emphasis is applications will be on fitting the transformation to be used with the job to be done.

7/65–203 **Information Systems for City Management.** October 1963. Edward F. R. Hearle. RAND Corp., Santa Monica, Calif. Rept. no. P2803 (AD–420 212). 11 pp.

7/65–204 **Set Theoretic Models for Classification and Retrieval.** November 1964. Richard Jernigan and Alfred G. Dale. Report LRC–64–WTM–5 of the Linguistics Research Center. University of Texas, Austin, Texas. NSF Grant GN–308. 20 pp.

The appropriateness of lattice models as analogues of classification and retrieval systems is analyzed. Models considered include simple distributive lattices, and implicative and subtractive lattices. A particular topological model is also considered. It is shown that any set of topological axioms will be reducible to a Boolean algebra when applied to finite spaces. (authors)

7/65–205 **Mathematical Formulation of Basic Procedures in Documentation.** April 13, 1960. James W. Perry and William Goffman. Western Reserve University, Cleveland, Ohio. Contract AF49 638 357 AFOSR TN60 366 (AD–429 098). 50 pp.; OTS price $4.60.

Certain basic aspects of documentation, especially the relationship of a *message* to alternate means for its expression and recording were studied. Attention was also directed to various documentation processes in which the message remains invariant, though the symbolism for expressing it may be greatly altered. In considering mathematical formulation, this led to such concepts as the message as a set, which may find expression in an equivalence class of sets, each of which is a version of a given message. The concept of set, in turn, led us to such concepts as sets of sets, sub-sets, and irreducible sub-sets—also sometimes called *unit elements.* The concepts of mapping and inverse mapping were also shown to be directly involved when dealing with such equivalence classes.

7/65–206 **Study of Theories and Models of Information Storage and Retrieval. Report No. 8: The Structure of Document Relations.** August 25, 1964. Donald J. Hillman. Center for the Information Sciences. Lehigh University, Bethlehem, Pa. NSF Grant GN–283. 29 pp.

A theory is proposed for articulating the relationships between documents *via* the relationships holding between the terms used to

characterize the documents. A basic predicate of connectedness is introduced, in terms of which term-term and term-document relationships are describable. Properties of connectedness are set forth in a number of theorems. A measure is then proposed for computing the strength or width within a document of a term used to characterize it, and it is shown that the formal system of connectedness has a graph-theoretical formulation. Each graph is relative to a prescribed width of index terms and a given value of matching between terms. On the basis of such graphs, a document set may be given a topology, rendering possible the performance of those retrieval operations for non-static document collections described in Report No. 6 of this Series. (author)

7/65–207 **How to Plot a Breakthrough.** December 12, 1963. Lauren B. Doyle. System Development Corp., Santa Monica, Calif. Rept. no. SP1492 (AD–427 161). 21 pp.

Will there be a breakthrough in the field of information retrieval? One authority in that field has said, "No." This paper adopts the opposite viewpoint, and speculates on what the elements of such a breakthrough might be if it were to occur. Several breakthroughs in other fields are scrutinized in order to highlight the factors which characterize and energize sudden expansions of new technologies. These factors, plus some factors specific to the field of information retrieval, are then extrapolated into a "plot for a breakthrough." (author)

7/65–208 **Documentation and Information Retrieval Aspects of Army Studies. Volume II, Annex C to the Army Study System. Study Documentation and Information Retrieval.** March 1963. Charles J. Davis. Department of the Army, Washington, D. C. (AD–602 527). 256 pp.; OTS prices: HC $6.00/MF $1.25.

The documentation and information retrieval aspects of Army study effort is examined. Eight relatively independent study programs are described. These programs are found to be documented by some seventeen different information flow mechanisms. The identified information flow mechanisms and the manner in which the programs are managed constitutes a weak *system* by documentation and information flow standards. A review of the *state of the art* of library science, documentation and information retrieval trends and experience is made to identify a range of alternative improvements. A modest pilot system, to include a central reference library with a searching aid and a periodic bibliographic publication, is proposed. Alternative follow-on improvements are cited as candidates for further research. (author)

7/65–209 **Cataloguing and Retrieval of Environmental Information: A Statement of the Problem.** April 20, 1964. Maurice H. Simpson. Army Frankford Arsenal, Philadelphia, Pa. Proj. 1A025001A62 19 07 AFA R1713 (AD–600 092). 32 pp.; OTS price $1.00.

The report describes the extent of the environmental information retrieval, storage and dissemination problem. It discusses the interrelated elements of the problem of general information needs and evaluation of natural and induced environments, and the specific needs concerning the effects of environmental stresses as related to when, where, and how equipment is to be used, transported, stored and handled. Added to the discussion are the need for standardization of environmental terminology, test methods, and procedures; reliability, failure analyses and modes of failure; and the problems of quantity and quality of input information. Desirable criteria for a usable environmental information retrieval, storage and dissemination system are developed in the report. (author)

7/65–210 **Text Reporting and Editing Device.** Final rept. April 1962. William R. Nugent. Inforonics Inc., Maynard, Mass. Contract AF30 602 3088, Proj. 4594, Task 459402 RADC TDR64 31 (AD–438 417). 32 pp.; OTS price $3.60.

An experimental model of an automated aid to editing of textual matter in machinable form has been designed and constructed. The principal purpose of this work is to demonstrate the usefulness and need for small scale, inexpensive editing devices. Off-line special editing devices can provide a high percentage of computer capability in editing for a low percentage of computer cost due to the comparatively lower requirement of complexity and speed. (author)

7/65–211 **Compiling a Technical Thesaurus.** January 14, 1963. T. L. Gillum. Defense Documentation Center, Alexandria, Va. (AD–420 504). 13 pp.

A thesaurus is defined as a device for controlling and displaying an indexing vocabulary. The vocabulary is controlled in the sense that the individual terms are carefully chosen and appear as distinct, though not inseparable, entities; it is displayed in such a way as to provide ready access from a given entry to related words that may be needed to index a paper or phase a search question. Some factors that may be considered in formulating the vocabulary for the thesaurus are discussed in terms of experience in compiling The Thesaurus of ASTIA Descriptors. Among the basic criteria to be examined are: 1. the volume and scope; 2. the technical competence of indexers; 3. resources, including time, money, and equipment; 4. the demands and speed of service that must be satisfied by the system. Problems that must be dealt with in evaluating the individual terms include synonyms, homographs, generic relationships, logical relationships and proper degree of specificity. (author)

7/65–212 **Document Selection Methods—A General Analysis.** 1963. James W. Perry. Arizona University, Tucson, Arizona. Grants AF AFOSR61 79, AF AFOSR62 289 AFOSR 64 0562 (AD–434 403). 64 pp. Reprint from Readings in Special Librarianship, ed. by Harold S. Sharp, Scarecrow Press, New York, 1963, pp. 507–573. (Copies supplied by DDC.)

This paper shows how traditional indexing and classifying methods relate to such recently developed techniques as uniterm computer searching, and the like. First discussing the *information explosion* as the natural consequence of increased emphasis on research and development during the past few decades, it then analyzes the performance and requirements of workable deep-indexing information retrieval systems. Emphasis is on the application of logic to specify the selection of pertinent documents from a universe of pertinent and nonpertinent documents rather than on the hardware which is employed in such selection. (author)

7/65–213 **System Components Information Center.** September 1963. Marvin G. Toll. Aeronautical Systems Div. Air Force Systems Command, Wright-Patterson AFB, Ohio. ASD TDR63 695 (AD–425 581). 26 pp.; OTS price $1.00.

This report describes a proposed information transfer program for the selection and approval of components used in Air Force systems. It is intended for use by designers and also by the management group approving what the designers recommend. A method is presented showing how technical characteristics are used as the basis for item selection. When a complete characteristic description is determined, an established index code substitutes numbers for the particular words thus permitting rapid machine search and retrieval. The program employs 1,000-ft microfilm facilities for storage and retrieval. It features technical summaries of engineering data as the output for quick determination as to whether the selected items satisfy the requirement. (author)

7/65–214 **An Evaluation of Data Collections for Plastics.** May 1964. Gunther Cohn and Carol Carr. Laboratories for Research and Development, Franklin Institute, Philadelphia, Pa. Contract DA36 034 501AMC0119A (AD–602 085). 50 pp.; OTS price $1.25.

Recommendations are made for a data system suitable for the Plastics Technical Evaluation Center on the basis of a study of various aspects of data and existing data systems. PLASTEC needs data to carry out its function as an information center and to answer requests. The data sources available include unpublished information, the report, and open literature in plastics. Data exists in tabular and graphic form and may be critical or unevaluated; provisions must be made for handling all aspects of data.

Methods for handling data range from manual to machine and information can be filed by material, property or by a random access method. The existing data systems provide examples of the various methods available for handling data. Although some existing systems handle plastics data, the coverage is not complete and there still remains a need for a unified data center in plastics. A manual system consisting of data sheets filed by material is recommended for the present. This system can easily be converted for machine storage and retrieval if necessary. (author)

7/65–215 **A Method for Improving Comprehension of Technical Communications through Statistical Testing and Information Retrieval Procedures.** 1964. Terrence Eugene Hobbs. Pittsburgh University, Pittsburgh, Pa. (AD–438 662). 63 pp.; OTS price $6.60. Master's thesis.

When words, symbols and figures are committed to writing, it is generally done with the intent of transmitting the writer's thoughts to those of others. The success of this intent is contingent upon a multitude of variables. This paper examines a few of the variables which determine techniques for transmitting information as applied to the specific field of technical writing. Attention is directed to the communication of specific ideas, through writing, related to technical type subjects. (author)

7/65–216 **Indexing and Abstracting Experimentation Support.** Final rept. January 31, 1963. Owen Roberts. Utica College of Syracuse University, Utica, N. Y. Contract AF30 602 2586, Proj. NR4594, Task NR459401 RADC TDR63 61 (AD–420 969). 64 pp.; OTS price $1.75.

The purpose of this effort was to provide the Information Storage and Retrieval Section (RAWIP–2) with the capability to investigate under direct RADC supervision several problem areas in the field of Information Storage and Retrieval. Each effort investigated under this contract had a direct relation to a contractual or in-the-house effort conducted by the Information Storage and Retrieval Section. (author)

7/65–217 **Applied Research Program Aerospace Intelligence Data System (AIDS). Volume II. Related Research Papers.** Rept. no. 6. November 16, 1962–February 15, 1963. 1 vol. Contract AF19 626 10 (AD–419 908). OTS price $6.60.

Contents: Toward Document Retrieval Theory; Techniques for Document Retrieval Research; On Natural Information Systems; and Evaluation of Clusters on the Basis of Random Graph Theory.

7/65–218 **Machine Recording of Textual Information During the Publication of Scientific Journals.** December 16, 1963. Lawrence F. Buckland. Inforonics, Inc., Maynard, Mass. 30 pp. (see also 1/65–202)

A program of research was undertaken to develop publishing and computer processing techniques for recording textual data in machine form at the time of primary journal publication. The purpose of recording journal data in machine form is so that it can be processed with the aid of machines to form secondary products, such as indexes and abstract journals. The approach to this objective consisted of two steps. The first, and most important, step of the project was the development of a system for recording journal articles so that text items were identified according to their functional end use. Next, transformation procedures were developed to convert the recorded data to the forms of the primary journal, indexes, abstract publications, and a data file. (author)

7/65–219 **The Grammar of Subject Headings: A Formulation of Rules for Subject Heading Based on a Syntactical and Morphological Analysis of the Library of Congress List.** Jay Elwood Daily. D.L.S., Columbia University. 1957.

Analyzes subject headings in Library of Congress list (5th edition, 1948) to determine function of each grammatical form in the structure of the list. Groups the 21,451 main headings according to form, punctuation, and inflections of words.

Finds that 47% of the main headings are composed of 2 or more words, while 31.5% consist of 1 word. Although there are significant variations in use of certain grammatical forms with suggested Library of Congress classifications, the pattern of usage suggests that study of a sample of headings based on a subject area would not lead to valid conclusions.

Comparison of grammatical forms used in seven special lists indicates that a single area of usage could be determined for each mark of punctuation, and rules for use of grammatical form are devised. The problems of grouping and dispersal of headings suggest need for a classified guide to supplement the alphabetic list of headings and the use of a mark of punctuation instead of prepositions and conjunctions.

7/65–220 **Correlative Indexing Systems for the Control of Research Records.** 1960. Gerald Jahoda. D.L.S., Columbia University.

Studies installations making use of correlative indexes in the physical and biological sciences to determine whether another sort of index (classified, alphabetic subject, alphabetico-classified, or another type of correlative) might have been employed with equal success in these situations. Sources of information include literature published in the field since 1948, questionnaires, and follow-up interviews with users of correlative indexes.

Concludes that "traditional" indexes can be used as efficiently as correlative indexes, except in area of chemical structures, and that, in general, different types of correlative indexes can be used with same degree of efficiency.

7/65–221 **An Investigation of the Techniques and Concepts of Information Retrieval.** Quarterly rept. no. 4. April 1-June 30, 1962. Jacques Harlow. ITT Federal Electric Corp., Paramus, N. J. July 31, 1963. Technical rept. 5201TR0058 Contract DA36 039SC90787 (AD–428 993). 69 pp.

Work in two of the four areas of capability described in the project task structure has been performed. Under input capabilities an extension of work on procedures for automatic assignment and the development of a probabilistic non-Boolean retrieval model has been continued. The work on non-Boolean retrieval is also relevant to the query capabilities subtasks concerned with the specification of scope and the relaxation of limitations on descriptions. Under processing capabilities there is no new progress to be reported on associative procedures, but extensive mathematical analysis has been continued on the problems of file organization and search. The analysis was extended to Markov chains. Under integrating capabilities the implicit interrelationships among the specialized studies have been reviewed, although there was no attempt to develop specific integrative measures; this work has been postponed in favor of extending the concepts in the specialized areas. (author)

7/65–222 **An Investigation of the Techniques and Concepts of Information Retrieval.** Quarterly rept. no. 5. July 1-September 31, 1963. Jacques Harlow. ITT Federal Electric Corp., Paramus, N. J. September 31, 1963. TR5201TR0069 Contract DA36 039SC90787 (AD–429 837). 47 pp.

Documentation in two of the four areas of capability described in the project task structure was produced in the last quarter. Under input capabilities a plan for the empirical evaluation on procedures for automatic assignment was developed. The economics of descriptor usage, importance of ranking of descriptors and automatic corrective procedures were considered. Work in the latter areas is also considered a significant contribution to ultimate system integration. Query capabilities were considered from the standpoint of a fact retrieval system and the problem of developing a logic of questioning is discussed. These considerations provide a framework for the requirements of further developments in processing capabilities. (author)

7/65–223 **An Investigation of the Techniques and Concepts of Information Retrieval.** Quarterly rept. no. 6. October 1-December 31, 1963. Jacques Harlow. ITT Data and Information Systems Div., Paramus, N. J. January 31, 1964. TR5201TR0078 Contract DA36 039SC90787 (AD–437 924). 91 pp.; OTS price $8.60.

This report discusses the work performed on the project to investigate the techniques and concepts of information retrieval and to formulate and develop a general theory of information retrieval. The formalization of this theory is oriented to the automation of large-capacity information storage and retrieval systems. This theoretical framework will be the basis for the use of general purpose stored-program digital computer systems to perform the storage and retrieval functions. (author)

7/65–224 **An Investigation of the Techniques and Concepts of Information Retrieval.** Quarterly rept. no. 7. January 1–March 31, 1964. Jacques Harlow and Paul W. Abrahams. ITT Data and Information Systems Div., Paramus, N. J. April 30, 1964. Rept. no. TR5201TR0088 Contract DA36 039sc90787 (AD–447 446). 48 pp.

A reformulation of the project tasks took place. Four questions were developed that encompass the major aspects of information retrieval system design. This reformulation accounted for the static as well as the dynamic aspects of information retrieval systems, and provided a means of organizing the material developed during the course of the project into a coherent whole and furthering the aims of system integration. Work was also performed on developing means of empirically verifying certain methods of automatic document classification based on the occurrence of clue words in the documents. (author)

7/65–225 **Switching Functions for Simplified Data Retrieval and Display Devices.** September 1963. J. A. O'Brien. MITRE Corp., Bedford, Mass. Rept. no. W6160 Contract AF19 628 2390, Proj. 481 ESD TDR63 442 (AD–424 796). 25 pp.; OTS price $2.60.

Equivalence or inequality functions from the comparison logic of retrieval devices and require logical complements to be processed. Coding rules are described which permit subsets of these functions (which do not require complements) to act as the original functions. Savings in cost up to 50% may be realized. Electronic, optical, and mechanical forms of these concepts are possible. In particular, words of an associative or content-addressed memory may be organized into fields where the number of bits to encode each field is minimum. A display device which uses a deck of punched cards between a light source and a human observer to simulate combinational circuits with manual switch input and indicator output is described. Light from cells and fields of cells may be interpreted to correspond to the response of the first and second state respectively of circuits using AND, OR, NAND or NOR gates. This concept may be generalized to a class of devices where many parallel planes control the flow of electromagnetic waves through aligned apertures. (author)

7/65–226 **A System of Data Banking and Retrieval for Educational Research.** 1960. Donald David Denum. University of Texas.

Attempts to develop a system for storage and retrieval of information for use by teachers and researchers. Investigates two types of systems based on concept of the bibliography of ideas: One provides for storing of document characteristics according to a prearranged pattern fixed at time of storage; the other (based on principle of coordinate indexing) combines flexibility of category manipulation with simplicity of classification, but contains some serious weaknesses.

Describes a modified system designed to minimize these weaknesses and to develop full potential of the coordinate indexing system. Sets forth reasons for, and a description of, various operational techniques, using both a large-scale random-access memory (the I.B.M. Ramac 305) and a simple searching machine (the sorter).

7/65–227 **The History of the Printed Book Catalogues in the United States.** 1960. James Ranz. University of Illinois.

Presents chronological history of the printed book catalog, from colonial times to the late 19th century. Explains why this form of catalog was so long regarded as ideal and why librarians were finally forced to abandon it in favor of the card catalog. Traces evolution of cataloging practices, considering matters of form and rules of entry, catalog arrangement, and various efforts to develop a scheme of centralized or cooperative cataloging. Data were secured from more than 1,000 printed book catalogs and from annual reports and histories of individual libraries.

7/65–228 **A Stochastic Theory of Documentation Systems.** 1960. Robert Edmond Booth and Harrison Morton Wadsworth, Jr. Western Reserve University.

Develops a number of statistical and mathematical models for various aspects of documentation, as guidelines for additional research. Study is based on literature and on the authors' background knowledge.

Describes and discusses possible choices of action in various documentation situations, taking into account recent developments in statistical decision theory. As an example, the theory of games developed by Van Neuman is employed to solve a problem involving the documentalist's decision in a searching situation. Playing the theoretical game evolved for solution of this problem are Nature and the documentalist. Nature's strategies are restricted to a finite test, and the documentalist uses Bayes' solution to evaluate alternative decisions.

The authors present the thesis that documentation consists primarily of a scientific attitude toward research librarianship.

7/65–229 **Role of the Technical Library in Support of an Information Center.** November 1964. M. Bloomfield. Hughes Aircraft Co., Culver City, Calif. (TM–804. AD–609 825.) 21 pp.; OTS Prices: HC $1.00/MF $.50. (Also pub. as EPIC–S–5, AD–608 377.)

The information center is defined as that organization which acquires, stores, indexes, analyzes and synthesizes data and information. The library acquires, indexes, stores and disseminates documents and information. The information center's major product appears in reports which have evaluated, analyzed, integrated and synthesised the data of a special scientific topic and presented it in graphic or tabular form. The library supports the activities of the information center by providing for its acquisitions. The library can also provide assistance to the information center through its familiarity with information retrieval principles. The information center can rely on the library to provide it with indexing and abstracting tools which provide a means of access to related information. The library also supports the information center with its current awareness announcement bulletins. (author)

7/65–230 **Information and Scientific Creativity.** June 1964. Calvin W. Taylor. Utah University, Salt Lake City, Utah. (Grant AF AFOSR144 63. AFSOR 64 2502. AD–609 486.) 20 pp.; OTS Prices: HC $1.00/MF $.50.

Paper presented at the Second Symposium of the Federal Council for Science and Technology, 13 Apr 64, on the topic, 'Technical Information and the Federal Laboratory.'

Descriptors: (*Scientific personnel, Performances (Human))), (*Information retrieval, Scientific personnel), Human engineering, Supervisory personnel, Symposia, Intelligence tests, Learning, Memory, Reasoning, Industrial psychology.

In this paper the problem of studying what constitutes effectiveness and creativity in a scientist is discussed. The way the scientist receives and handles information, the intellectual climate in which he works, and the nature of the information received by him are all examined in their bearing on the creative process.

7/65–231 **An Announcement-and-Request System for Initial Dissemination.** Roy Davison and John Irvine. *Parameters of Information Science.* pp. 111–115. (see 4/65–1)

For selective initial dissemination among persons in real-time working relationships, System Development Corporation employs an announcement-and-request system, using potential readers as the selecting mechanism. Time from printing to reader's desk: as low as 24 hours. This system appears justified by its service, quality (the relevance/dissemination ratio) and cost. It handles from 20 to 150 (average 50) new titles each day, disseminating an average 600 copies daily among 2,500 readers. (authors)

7/65–232 **Uses of the Information Sciences from the Viewpoint of a Practicing Documentalist.** Lea M. Bohnert. *Parameters of Information Science.* pp. 49–51. (see 4/65–1)

A narrow definition of the Information Sciences, excluding psychology, will be adopted in order to show applications to the training of documentalists and to evaluation of work in the field. Symbolic logic and communication theory are shown to furnish illuminating guidelines for documentalists. Methods available to documentalists to accomplish their jobs are shown to be reducible to either *marking* or *parking,* to use Fairthorne's terminology. (author)

7/65–233 **IR Systems for Schools of the Future as Extrapolated from Current Research.** Don D. Bushell. *Parameters of Information Science.* pp. 53–56. (see 4/65–1)

Uses of a dynamic system of information retrieval in the schools or universities of the 1970's are summarized. Details of prospective uses of such an IR system by the student and by the teacher, researcher, or counselor are given, including the very promising use of on line information retrieval by the teacher in the classroom for both instruction and classroom management purposes. Several research projects in the areas of IR, language processing, and computer-based programmed learning are reviewed. (author)

7/65–234 **Information in Decision Making.** C. J. R. Licklider, *Parameters of Information Science.* pp. 9–21. (see 4/65–1)

Major decisions stem from interactions among assumptions, data, principles, heuristics, subordinate decisions, and value structures. This paper examines those interactions, analyzes their requirements and shortcomings, and proposes solutions based on improved man-computer communications. (author)

7/65–235 **On Concept—Formation and Relevance.** Donald J. Hillman. *Parameters of Information Science.* pp. 23–29. (see 4/65–1)

The foundations of a mathematical theory of relevance are described, and a connection established with a previously constructed topological theory of retrieval. A basic matching predicate is given a concrete interpretation via an index term association matrix developed by Giuliano and Jones, and the graph-theoretic properties of matching are set forth. On the basis of the interpreted predicate, index terms are sorted into genera, i.e., groups of conceptually related items, all orderings within genera are described, and concepts themselves are defined. It is shown that a document corpus can be given a topology via a topology of terms, and retrieval operations for non-static collections are briefly discussed. (author)

7/65–236 **Quality Control of Coordinate Indexing.** D. W. King and J. M. Daley. *Parameters of Information Science.* pp. 389–392. (see 4/65–1)

A number of retrieval studies has shown that indexing errors may be large and may have a considerable effect on the performance of the system. Therefore, a means for controlling indexing errors is desirable. Previous studies has discussed consistency and accuracy of indexing, with examples and tables. (authors)

7/65–237 **The Use of Consoles in Machine Translation.** Gerhard Reitz. *Parameters of Information Science.* pp. 277–284. (see 4/65–1)

The principal application of consoles is in post-editing of MT: choosing one of several English equivalents, changing and supplying missing words, rearranging word order. When debugging, programmers can correct instructions. Checking solution techniques, linguists can force the program flow at various levels of branchpoints. (author)

7/65–238 **Deep Subject Indexing by Manual Permutation Methods.** Saul Herner, *ADI Proc. Pt. 3.* (see 4/65–1)

Originally aimed at the permutation of textual items derived from various parts of publications, permutation indexing evolved in practice as a means of permuting titles of articles and other parts of bibliographic citations. Recently, there has been a revived interest in the permutation of textual items other than bibliographic citations. This paper discusses several examples of this more general type of permutation, and describes a procedure for permuting selected portions of texts by manual means to produce permutation indexes having the appearance of conventional book indexes. (author)

7/65–239 **New Role for Microfilm Proved Aboard University of Seven Seas.** Charles A. Koepke, III. NAMN. no. 72, October 1964. pp. 57–61.

7/65–240 **Legal Considerations for Microfilming in the State of Utah.** T. Harold Jacobsen. NAMN. no. 72, October 1964. pp. 62–9.

7/65–241 **Technickoekonomicka Propaganda v Pruomyslovych Zavodech *(Dissemination of Technical Information in Industry).*** J. Karlichek. Publ. Prace, Prague, 1964. 101 pp. Reviewed by J. Podzimek, in *Technicka Knihovna (Technical Library),* no. 9, 1964. pp. 279–280. In Czech.

The book is a guide for information specialists in Czechoslovakian industry. It describes the cooperation between various departments of an industrial enterprise; the duties and responsibilities of the information centers and information specialists; methods for the dissemination of technical information; planning information services; and effects of information services on the productivity of the enterprise. The organization of technical information in the Automobile Factory J. Dimitrov in Letnany is reported as a practical example of systems application. (Doc. Inc.)

7/65–242 **New Work Methods in Technical Libraries.** A. Vejsova, PhD., *Technicka Knihovna (Technical Library),* no. 2, 1964. pp. 25–40. In Czech.

The article discusses library mechanization and automation aimed at improving the processing of new technological and scientific information on a large scale. Cooperative acquisition on a limited scale, founded on subject area delimitation stipulated by library plans, was realized in Czechoslovakia in 1963. The basis for further developments in cooperative acquisition are union catalogs (local, regional and national). The paper evaluates the parameters of various catalogs from the viewpoint of employing mechanized methods of catalog compilation. Mechanized systems used for this purpose in the United States are also discussed selectively in some detail, including systems using tabulators, computers, and photographic methods. (M.H.G.)

7/65–243 **Probability and the Library Problem.** M. E. Maron, *Behav. Sci.,* vol. 8, no. 3, July 1963, pp. 250–256.

This is an abbreviated version of a publication by Maron and Kuhns which appeared in *J. ACM* 7 (1960), 216–244. That paper is recommended as a more adequate consideration of probabilistic indexing. Perhaps this second time around is justified inasmuch as the empirical studies required for an evaluation of weighted tags and relevance number have not yet appeared, at least to the knowledge of this reviewer. The concept of ordering search output by relevance number is very appealing. On the other hand, this probabilistic approach to indexing adds to the already heavy responsibility of the human indexer.

7/65–244 **Engineering and Scientific Documentation.** James W. Perry, Arizona U., Tucson. 1963, Grant AF AFOSR61 79, AFOSR J1145. 5 pp. Reprint from Indian and Eastern Engineer, no. 104, pp. 215–220. (Copies supplied by DDC.)

7/65–245 **A Comparison of Gothic Elite and Standard Elite Type Faces.** J. G. Fox. ERGO vol. 6, no. 12. April 1963. pp. 193–198.

Two forms of type faces were compared for readability, when the emphasis was on comprehension and when it was on

speed. No significant difference was found between the reading times for the two type faces when reading for comprehension but Standard proved superior to Gothic when speed was the more important. Possible reasons for these differing results under the two conditions are considered. (author)

7/65–246 **An Analysis of the Designing, Installation, and Operation of a Coordinate Indexing System Using Links and Roles for the Plastics Department of the Du Pont Company.** Barbara A. Montague. JCHD, vol. 4, no. 4, October 1964, pp. 251–255.

Discusses the decisions involved in the planning and operation of a technical information system incorporating internal research reports, patents, and sales reports dealing with polymers and polymer intermediates.

7/65–247 **The Technical Information Service in the Stanford University Libraries.** Jack Pooler and David C. Weber. *College & Research Libraries*, vol. 25, no. 5. September 1964, pp. 393–399.

Technical Information Service, a department of Stanford University Libraries which lends library materials and supplies photocopies and other services to industrial firms, is described. The relationship between the university library and the industrial library is discussed. TIS's history since its organization in late 1958 is presented, with detailed comments on staffing, quarters, charges for service, and budget. The future of the university library's relationship with government and industry is given consideration. (authors)

7/65–248 **Identifying Key Contributions to Information Science.** October 1964. Carlos A. Cuadra. AMDO, vol. 15, no. 4, pp. 289–295. Based on and supersedes report SP–1467 of the system Development Corp., Santa Monica, Calif. (AD–428 595). (see 4/64–9)

7/249 **Systematizing the Punch-Card Methods.** In German. Zur Systematik der Lochkartenverfahren.-Hans-Peter Giercke. NADO, vol. 15, no. 2, June 1964, pp. 82–85.

Systematizes the requirements concerning documentation. A literature card file serves as an example in using the system studied. Detailed reasons are given for the necessity of using several card files.

7/65–250 **Computers and Classification Systems.** Robert R. Freeman. *J. Doc.*, vol. 20, no. 3. September 1964. pp. 137–145.

Published experiences dealing with the use of data processing equipment in conjunction with classificaton systems, particularly the UDC, are reviewed. Earlier conclusions, generalized from experience with punched card equipment, were too pessimistic. Difficulties in handling UDC notation in computer systems and methods for overcoming them are discussed. Ways in which computers can aid in studying, improving, and using classification systems are discussed. Experience at Meteorological and Geoastrophysical Abstracts is cited. (author)

7/65–251 **Pre-Convention Microfiche Session.** NAMN. No. 71. August 1964. pp. 8–20.

At the request of four agencies of the Federal Government, notably the Atomic Energy Commission, the Department of Commerce (Office of Technical Services), the Department of Defense and the National Aeronautics and Space Administration, a pre-Convention panel discussion was held in Philadelphia on April 27 immediately prior to the opening of the 13th NMA Convention. The purpose of the meeting was to announce the adoption by the four agencies of a uniform 105 x 148 mm. microfiche size for technical reports and to explore the implications of the use of a single standard size format in terms of equipment design and the use of information.

7/65–252 **On Automatic Documentation and Selection of Medical Findings—Their Technical and Medical Problems.** August 3, 1963. H. J. Heite, West Germany. *Medizinische Welt*, (31): 1560–1570. In German.

The modern hospital, despite its progress in other fields, has net yet adopted the present methods of data processing for its medical records. The various procedures for information storage and retrieval, as they are used in modern business organizations, are described and their advantages and limitations in the vast field of medical records shown. The basic unit is the IBM punch card whose fruitful use depends on the presence of certain technical facilities in the hospital, e.g., a sufficient number of machines producing perforated tape. Does the punch card actually satisfy the requirements the physician and scientist expect of automatic documentation and selection of medical record material? An ideal documentation machine should store the entire content of a patient's record for future retrieval at any time. But the actual punch card, as available today, furnishes only documentation restricted to a multifold data selection. All kinds of findings are compiled and summarized in a rather restricted number of classification units. The writing of the whole medical card consists of nothing but the arrangement of the given data into preconceived classes and headings. Instead of providing an individual analysis of a medical record in all its various details one merely chooses among a restricted number of prepared answers the one which fits the case best. This disadvantage can be overcome by magnetic tape systems which furnish a codeless scanning procedure. But this new system has not yet been used in Germany for the purpose of medical record documentation. Until this procedure has been sufficiently tried out, automatic data processing can only be used for specific scientific research. (H.W.B.)

7/65–253 **Clippings and Possibilities for Their Use in Technical Libraries and Technical and Economic Information Departments.** 1964. Podzimek. *Technicka Knihovna.* (Technical Library), no. 1, pp. 8–10. In Czech.

A clipping service, effectively complementing information available through classical library methods, is the simplest form of building up a complex fund of information. It is also more economical than microfilm and photocopies.

The article discusses the activities and extent of the services rendered by the clipping service of the Prague Information Service, the nationwide center for the processing of clippings. Clippings are recommended as a supplement to specialized literature; they can supply information available on any given topic in print, both domestic and foreign. The clippings are selected to meet the subscriber's needs, and may be used as a basis for his future activities.

The Czechoslovak National Clipping Service [address: Vystrizkova Sluzba Prazska informachi sluzby, Prague 2-Vysehrad, K rotunde 8/82] covers all daily Czechoslovak press, including local district and factory publications; the press of the countries of the socialist bloc; and the main "progressive" press of the Western world. (P. Z.)

7/65–254 **Categories and Relations: Origins of Two Classification Theories.** R. Moss. AMDO Vol. 15, No. 4, October 1964. pp. 296–301.

The resemblances between the Categories of Aristotle and those of Ranganathan are shown. These categories are examined in the light of criticism made by Bertrand Russell and are shown to have no validity. Similar comparisons are made between the Relations of Hume and Farradane. Farradane's work is a return to Hume, who is generally acknowledged as one of the founders of the British school of empirical philosophy which continues to Russell and beyond. In Russell's work lies the most promising line of development for information classification and indexing. (author)

7/65–255 **Theoretical Considerations in Information Retrieval Systems.** Jack Belzer and W. Goffman. CACM, Vol. 7, no. 7, July 1964, p. 439.

Information storage and retrieval systems are composed of three maor components: (a) identification of information and tagging it for effective retrieval, (b) searching strategy, how to enter the file to circumvent the scanning of nonrelevant material, and (c) file organization to make access to information efficient. For identification of information the paper suggests that a meta-language (recently discussed in a paper by Goffman, Verhoeff and

Belzer) associated with an object language be used. For searching strategy, a linear model for an evaluation function of relevancy is developed which rewards the system for retrieving relevant documents and not retrieving the nonrelevant, and penalizes the system for the escaped relevant documents and false drops. The inadequacies of a linear model are indicated. Two approaches to file organization are discussed. One is self-organization of the file based on its history and past performance, and the second is a self-generating subset of the file with a high probability of being relevant. (authors)

7/65–256 **Technical Film—The Most Effective Tool for the Dissemination of Technical-Economic Knowledge.** J. Stverak. *Technicka Praca (Technical Activity),* February 1964. pp. 147–148. In Czech.

The author evaluates the production, distribution, and use of technical films in Czechoslovakia. At present, films are commissioned by individual ministries or enterprises according to their needs. The studios are capable of producing one hundred films annually, but lack of planning has resulted in the delivery of only 80% of those ordered. The author suggests that the State Commission for Scientific and Technical Coordination and Development coordinates the planning and financing of new films with the national plan for scientific and technical development. Each film should deal with a single topic and be clear and understandable. Schools are especially interested in five minute films and filmstrips for use as teaching aids. Regional film lending libraries should actively promote the use of their holdings by co-operating with scientific organizations, trade unions, and schools. A proper catalogue of available films has has not been prepared. A documentation method should be worked out which would incorporate the films into the overall technical-economic information system. In 1963, an average of 4.67 films per 1,000 population were demonstrated in Bohemia and Moravia. In Slovakia, the average was 2.31 per thousand. (M.H.G.)

7/65–257 **Who Is Responsible for Communication?** Edward C. Walterscheid. IEEW, vol. EWS–7, no. 2, September 1964. pp. 8–10.

This paper assumes that such responsibility rests mainly with the editor and the writer and explores several viewpoints as to the primary responsibility, with examples made of government technical reports and of the *Proceedings of the IEEE.* (BATR)

7/65–258 **A Mid-Century Survey of Books on Communication.** Martin P. Anderson. Journal of Communication, vol. 14, no. 4, December 1964. pp. 203–214.

This paper presents the findings of a survey of books published during the period 1950–1959. It identifies significant books and five trends. (BATR)

7/65–259 **An Editor's View of the Past and Future Impact of Computer Technology on the Publication Explosion in Education and Psychology.** Fall 1964. William B. Michael. *J. of Educational Data Processing.* pp. 253–9.

7/65–260 **Origins of the NMA Microfiche Standard Specifications M-1-1963.** NAMN No. 66. October 1963. pp. 53–76.

7/65–261 **Automatic Information Processing in Western Europe.** Gerard Salton, SCIE Vol. 144. No. 3619. May 8, 1964. pp. 626–632.

7/65–262 **Planning for Scholarly Photocopying. A Report Prepared for the American Council of Learned Societies.** *Publications of the Modern Language Assoc. of America,* vol. 79, no. 4, pt 2. September 1964. pp. 1–14.

7/65–263 **MR&S Now Publishing Key Words and Abstracts.** MTRS, vol. 4, no. 1. January 1964, pp. 34–35.

7/65–264 **The Technical Literature in Continued Education.** July 1964, R. J. Schwarz. Proc. IEEE.

7/65–265 **A Communication: Organizing the Literature.** Gerald M. Shattuck. *International Development Review,* vol. 5, no. 3. September 1963, pp. 32–33.

7/65–266 **Proposing a Computerized Development Information System.** J. Ben Lieberman. *International Development Review,* vol. 6, no. 1. March 1964, pp. 33–35.

7/65–267 **Development Information: Do We Need Computers?** John Martinson. *International Development Review* vol. 6, no. 2. June 1964, pp. 35–36.

7/65–268 **Retrieval and Reproduction: Past and Future.** August 1964, Hugh W. States. NAMN, no. 71, pp. 2–7.

7/65–269 **Normes polonaises concernant les microfilms.** Janina Pelcowa, REDO, vol. 31, no. 3, August 1964, p. 111.

In 1962-1963 three Polish standards have appeared, non-existent abroad, which lay down rules for the form of catalogue entries of microfilms, and for the title description and imprint which microfilms should carry.

Details of the standards are described. (author)

7/65–270 **Experiments in mechanized control of meteorological and Geoastrophysical Literature and the UDC Schedules in These Fields.** Malcolm Rigby. REDO, vol. 31, no. 3. August 1964, pp. 103–106.

Described are experiments, carried out by *Meteorological and Geoastrophysical Abstracts,* in mechanized thesaurus manipulation and updating, current title arrangement and indexing, and bibliographic preparation, using punched card or magnetic tape input and UDC classification, as well as selective retrieval and printout of titles.

Other experiments concerned UDC tables: UDC schedule maintenance, updating and presentation, multi-lingual UDC manipulation and presentation, and alphabetical indexing of systematic or hierarchical classification schedules.

The UDC is considered competitive as a system for use in computer retrieval and printout systems.

Projects needing further exploration are listed. (author)

7/65–271 **Information, Please.** 1964. J. Richard Elliott, Jr. BRNS. October 19, pp. 3, 10, 12–13; November 2, pp. 3, 8, 13–17.

This semitechnical, economic survey of the IT (information technology) business outlines, with some minor inaccuracies (especially in technical nomenclature), most of the recent innovations in hardware and software devices for information handling In general, despite some technological breakthroughs, financial rewards so far have been less than spectacular, but certain companies begin to operate in the black. (C.L.B.)

7/65–272 **The Translating Unit (at N.I.H.).** Anonymous *Lab. Management,* vol. 2, no. 5, September–October 1964, pp. 28, 43

The organization and operation of the Translating Unit within the Readers Services Section of the N.I.H. Library are briefly described. One administrator, 7 full-time translators (covering ca. 15 European languages), and 3 editorial clerks handle the translation needs of 11,000 scientists, engineers, and administrators within N.I.H. German, French, and Russian are in greatest demand. (C.L.B.)

7/65–273 **Documentation Services to An Encyclopaedia.** G. J. Narayana. REDO, vol. 31, no. 3, August 1964, pp. 112–116.

The nature and functions of documentation services to an encyclopaedic compilation is described. Gives a detailed account of the documentation services, i.e. collection, evaluation, storage and retrieval of information for the encyclopaedic compilation *Wealth of India,* followed by a description of the building and organization of literature files. (author)

7/65–274 **"Citation Index" and "Ruckwartskatalogisierung." Instances of Citation Indexes.** (In German.) "Citation Index"

und "Ruckwartskatalogisierung." Beispiele für Zitatendokumentation. Fritz Model. NADO, vol. 15, no. 3. September 1964, pp. 122–130.

The *Science Citation Index* offers a fundamental documentation with cross-relations, the "Ruckwartskatalogisierung" is a compilation of the literature of a special field.

7/65–275 **Library Science and Documentation. vol. IV. Scientific and Technical Libraries. Their Organization and Administration.** 1964. Lucille J. Strauss, Irene M. Strieby, and Alberta L. Brown. 398 pp. Wiley & Sons, New York. (Z675.T3 St82s)

Provides an introduction to the organizational procedures and essential functions of a special library or information service in the subject fields of the sciences and their related technologies.

7/65–276 **Rapid Structure Searches via Permuted Chemical Line-Notations.** Peter F. Sorter, Charles E. Granito, John C. Gilmer, Alan Gelberg, and Edward A. Metcalf, JCHD vol. 4, no. 1. January 1964, pp. 56–60.

Study of permuted line notations has shown that tabulated lists can be used to locate rapidly specific compounds, classes of compounds having similar ring systems and all compounds having the same functional group. The need for a functional group field for a chemical structural retrieval program can be eliminated. This approach appears to be economically feasible for medium and possibly large files of structures.

7/65–277 **Documentation of Documentation. H. A. Comparison of Information Services in the Field of Documentation.** (In German.) Dokumentation der Dokumentation. II. Ein Vergleichvon Informationsdiensten uber Dokumentation. Herbert Buntrock. NADO vol. 15, no. 3. September 1964. pp. 145–147.

Describes some results of a study of information services in the field of documentation in regard to the analyzed periodicals, the number, time delay, and other important characteristics of abstracts and reviews, title lists, and bibliographical strips.

7/65–278 **The Research Librarian in a Challenging Age.** Walter M. Carlson, SPLB, vol. 55, no. 1. January 1964, pp. 11–19.

Describes briefly three priorities of the technical information program of the Department of Defense. Major trends in the field of technical information include the decline of the technical journal as a means of promptly conveying scientific and technical information; the addition of new terms and concepts to our technical vocabulary; more and more administrators are recognizing that communication of technical information is an integral part of the work; and scientists are finding it profitable to extract, evaluate results of work in their specialized fields.

7/65–279 **Classification and Coding.** J. Gombinski. *Engineering Materials and Design*, vol. 7. September 1964, pp. 600–605.

Discusses the relevance of classification to quantitative studies and organizational methods, with a particular emphasis on engineering. Discusses the principal fields of application of classification and coding which are: design, inventory, production planning, and data processing.

7/–65280 **Some new Concepts in Library Service.** W. K. Lowry. BLRC, vol. 42, January 1964, pp. 3–7.

Describes the activities of the Bell Laboratories libraries concerning their announcement bulletins, computer prepared publications, and the motion picture film service being developed.

7/65–281 **Research Methods in Librarianship.** Guy Garrison, Ed. LIBT, vol. 13, July 1964. 149 pp.

These papers, with the exception of the final one, were originally prepared for the Conference on Research Methods in Librarianship sponsored by the Library Research Center, Graduate School of Library Science, University of Illinois. September 1963.

7/65–282 **Advanages of a Combined Preparation and Use of Machine Punched Card, Slotted Card and Peek-A-Boo Card Files in Patent Documentation.** (In German.) Vorteile einer kombinierten Herstellung und Anwendung von Maschinen-, Schlitz- and Sichtlochkarteien bei der Patentdocumentation. Otto Schon. NADO, vol. 15, no. 3. September 1964, pp. 134–139.

For patent documentation a system combining several kinds of file cards is proposed.

7/65–283 **Science Citation Index.** Eugene Garfield. SCIE. vol. 144. May 8, 1964. pp. 649–654.

A new dimension in indexing.

7/65–284 **A Basic Theory of Roles as Syntactical Control Devices in Coordinate Indexes.** John C. Costello, Jr. JCHD. vol. 4, no. 2. April 1964, pp. 116–124.

Describes the EJC System of roles and gives examples in Latin and Finnish language.

7/65–285 **Sixth Annual Institute in Technical and Industrial Communications.** July 8–12, 1963. Colorado State University, Fort Collins. Reprinted from *BioScience*, vol. 14, no. 1, 1964. pp. 33–34.

Dr. Israel Light of the NIH Division of Research Grants discussed communication problems in the biological-medical sciences. He itemized the strategic elements of the communication process and developmental trends in his field and described problems accompanying this information explosion.

Dr. Dwight Gray, Chief, Science and Technology Division, Library of Congress, spoke on "Science Writers and Editors vs. Readers: What do the Former Owe the Latter?" He stated the principle of the Weinberg report and other documents that publication of research results is an integral part of research.

Dr. Joseph F. Montague, Director of the Medical Writers Institute, distinguished between "authors" (originators of information) and "writers" (disseminators). In medical terms he described communication as the transference and successful transplant of information in another mind. He noted that friction is developing between the academic and commercial writer in the medical sciences. Dr. Montague emphasized that each has his place and each must assume responsibility for his writing; that the "tradition" of both should be that of medicine.

Dr. J. W. Perry, University of Arizona, well-known for his work with computer handling of information, listed four aspects of communication as related to machine handling: (1) symbols, discourse, and signal transmission; (2) recording and remembering; (3) requirement of decision-making; and (4) perception.

7/65–286 **Challenges in microimage publishing.** 1964. O. A. Ullrich and Lewis E. Walkup. *TAGA Proc.*, pp. 248–278. (Rochester, N.Y., Tech. Assoc. of the Granphic Arts).

In addition to present archival uses, microforms could well supply the general public's everyday reading needs (books, periodicals, and even newspapers). The psychophysical reason for the lag in the exploitation of this vast potential is attributed primarily to manufacturers' disregard for the comfort and preferences of the users of microfrom viewers. Specifications for an acceptable reading device are presented, and an experimental model reader based on them is described and illustrated. (C.L.B.)

7/65–287 **UNESCO Meeting on Scientific Translation and Terminology.** R. Walter Jumpelt. *Babel.* vol. 10. no. 1. 1964. pp. 25–28.

Gives the recommendations from the meeting January 27–February 1, 1964 in Rome.

7/65–288 **Automation of Reference Work.** Claire K. Schultz. *Library Trends*, vol. 12. January 1964, pp. 413–424.

Describes the use of the peek-a-boo cards, and machine retrieval systems available for reference work. Search strategy and the MEDLARS system are discussed.

7/65–289 **The Technical Order Library in Orbit.** R. M. Winz. *STWP Review,* vol. 11, no. 1. January 1964, pp. 2–4.

Manned orbital space stations may present problems in providing technical order support. Three solutions presented are ground-station-to-space-station communication, orbiting data storage station, and on-board orders using a technique involving photochromic microimages.

7/65–290 **The Singularity Sub-Link. A New Tool for Use in the Storage and Retrieval of Information.** J. Frederic Walker. JCHD, vol. 4, no. 1. January 1964, pp. 45–48.

The singularity sub-link which gives additional scope and utility to concept coordination systems, has been successfully employed in a technical report indexing system in the Electrochemicals Department of the du Pont Co.

LITERATURE NOTES—Please enter my subscription for one year commencing with the January 1965 issue. At $4.00 per year, payment is enclosed ☐

bill me. ☐

Name__________ Title__________

Address ☐ Business ☐ Home__________

City__________ State__________ Zip__________

Your Firm Name__________

Literature Notes Separate

The Literature Notes section of American Documentation is now available as a separate publication.

Literature Notes will continue to be an important section of American Documentation. However, popular demand has persuaded ADI to offer this separate quarterly subscription as an offprint.

On a Calendar year basis, the subscription price of Literature Notes is $4.00 per year, commencing with the January 1965 issue.

Return the coupon above. Payment with the order is requested.

Attention: Director of Technical and Scientific Information

Dear Sir:

We are now delivering to scientific and technical information centers, for the first time on magnetic tapes, interdisciplinary files describing the current literature of science and technology. This information can be used in-house for many purposes. Among them are:

- Input for selective in-house alerting publications
- Selective Dissemination of Information services
- Input for KWIC, KWOC indexes, etc.
- Retrospective searching
  - Author Indexes
  - Title Indexes
  - Journal Indexes

Coverage and Contents of Data Bank: This file covers all articles, reviews, proceedings of meetings, technical notes, letters to the editor, book reviews, and all other items of permanent value published and to be published in more than 1,060 key scientific journals of 1965 and all current U.S. Patents issued including all primary patents and reissued patents.

Each item carries a unique identification number. For items published in journals, authorship is provided (with no limit on the number of authors) along with the titles of the items and basic citation information including volume, issue number, ISI accession number, starting page, number of references cited in the item and a code characterizing the type of item. When the entry is a patent, all inventors, co-inventors, and assignees are provided along with the patent number, date of issuance, number of references cited, Patent Office classification number, and complete patent titles.

> Complete computer programs for the IBM 1401/7074 are available without charge to purchasers of these magnetic tapes. These programs comprise a sophisticated system for the selective dissemination of information based on any Boolean Algebra combinations desired and include an adjustable weighting factor for each subject term.

If you would like to explore the possibilities of these tapes, or if you want additional information, please telephone collect.

Sincerely,

Philip Sopinsky

Philip Sopinsky, Manager
Data Processing Department
PS/s

Note: The SDI system described above was developed by the Ames Laboratory, Iowa, and is available to the public without charge.

INSTITUTE FOR SCIENTIFIC INFORMATION
325 Chestnut Street Philadelphia Pa 19106 / Telephone 215-923-3300

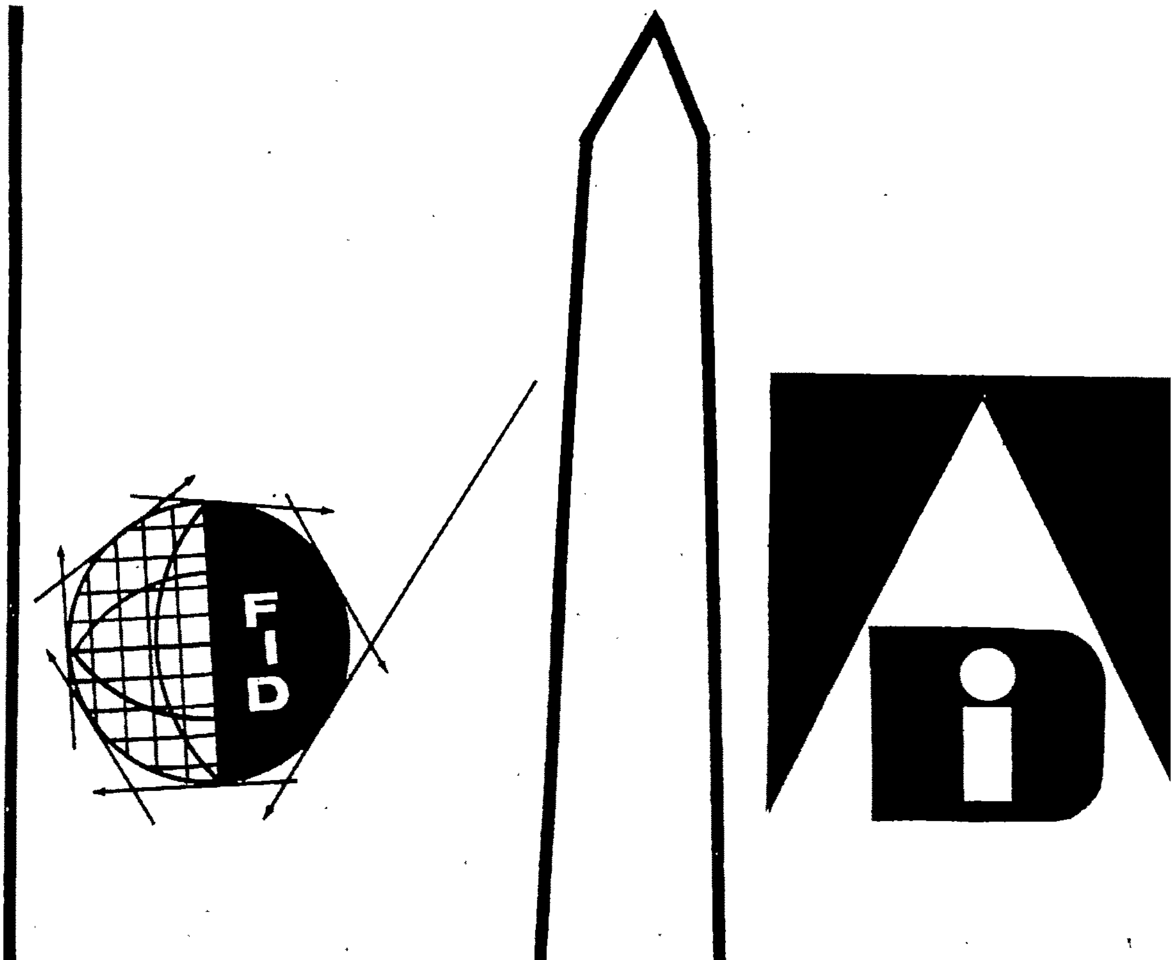
OCTOBER

1965

Vol. 16, No. 4

American Documentation

PUBLISHED QUARTERLY BY THE AMERICAN DOCUMENTATION INSTITUTE



31st MEETING AND CONGRESS
INTERNATIONAL FEDERATION FOR DOCUMENTATION (FID)
WASHINGTON, D.C. 7-16 OCTOBER 1965

AMERICAN DOCUMENTATION

American Documentation is a publication of the American Documentation Institute. Its purpose is to be a scholarly journal in the various fields of documentation and to serve as a forum of discussion and experimentation.

American Documentation is published in January, April, July, and October. One copy is included in the individual membership fee ($20.00 per year), one copy in the institutional membership fee ($100.00 per year), and up to five copies in the sustaining membership fee ($500.00 per year). Nonmembers may subscribe at $18.50 per year, postpaid in the U.S. Single copies may be purchased for $4.65 each. Communications concerning memberships, subscriptions, reprints, renewals, back issues, advertising, and changes of address should be sent to the American Documentation Institute, 2000 P Street, NW, Washington, D. C. 20036.

Instructions to authors are given in Volume 14, No. 2 (April 1963). All manuscripts, brief communications (1000 words or less), announcements, or letters should be sent to:

Mr. Arthur W. Elias, Editor
American Documentation
Institute for Scientific Information
325 Chestnut Street
Philadelphia, Pennsylvania 19106

Correspondence on reviews should be addressed to:
Dr. Theodore C. Hines, Review Editor
American Documentation
54 North Drive
East Brunswick, New Jersey

Correspondence on Literature Notes should be addressed to:
Mr. Burton E. Lamkin, Literature Notes Editor
American Documentation
1649 Fairorchard Ave.
San Jose, California 95125

American Documentation is indexed in *Library Literature, Current Contents of Space, Electronic & Physical Sciences, Library Abstracts, and Science Citation Index.*

American Documentation is entered for second class mailing at Washington, D. C.

work presented by the mathematical apparatus is much too vague to be of great help in describing the specific structures of particular languages or in processing these languages on a computer.

An additional feature of this approach is the development of further mathematical apparatus to describe in the same very general terms not only the object languages but a number of metalanguages deemed necessary to deal with the object languages. This additional aspect of the formational approach to grammar seems to be no more oriented to practical applications than the previously mentioned aspects.

Transformational Grammars. These are the best known and most popular of the formal approaches to grammar today. They arose as an alternative to the finite-state and phrase-structure grammars which are considered inadequate by the proponents of the transformational approach.

This theoretical approach is under constant development and new facets of the theory emerge almost daily. It is therefore difficult to give an evaluation of it in every detail, since it is not certain that a particular detail of the theory will still be accepted at the time the evaluation is made.

The claim made by the proponents of the transformational approach to grammar is that it has the "greatest explanatory power" of all grammatical theories known to date. This is a very strong claim and at the same time a very vague one, since the term "explanation" has no precise meaning in scientific parlance.

The essential feature of the original version of transformational grammar is that only certain basic types of sentences, called *kernel sentences,* can be accounted for by a phrase-structure approach. The remaining more complex sentences can be accounted for by deriving them from the kernel sentences by means of a set of special rules called transformation rules. These derived sentences are called *transforms.* The sentences of a language can then be classified as kernel sentences and transforms.

A more recent version of transformation theory has introduced modifications in the transformation rules and has introduced a new significant characterizing principle into formal grammar. This new principle is the distinction in the characterization of sentences between their "surface structure" and their "deep structure." The deep structure can be specified by means of a further refinement of a phrase-structure description. The surface structure is then revealed by the application of certain transformation rules to the deep structure. All of the information necessary for generating the surface structure is already contained in certain special markers that form part of the deep structure. These markers automatically trigger the transformations which bring about the surface structure.

Two concepts are of great interest in the transformational approach to grammar: one is the suggestion that certain sentences are derived from other sentences; the second is that the difference between simple and complex sentences corresponds to a profound difference in the underlying structure rather than merely the presence of a greater variety of detail. Both of these observations are significant for the design of syntactic recognition routines for machine translation. The transformational approach, however, is not intended to answer the fundamental question for practical application. Namely—What are the recognition criteria by which a computer program can correctly identify the boundaries and functions of the different components of a sentence?

Stratificational Grammar. This approach is an attempt at formalizing the concept that the structure of language consists of several hierarchic levels. This conception has been accepted among American descriptive linguists for some time.

The stratificational model proposes that natural language contains four main levels, called *strata.* There are, going from the lowest to the highest stratum: the phonemic, the morphemic, the lexemic, and the sememic. Each stratum has its own units and rules of combination. The relation between the different strata is expressed by rules of representation. The entities on a higher stratum are represented by entities on the next lower stratum—for instance, entities on the lexemic stratum by entities on the morphemic stratum, etc. The terminology and logical characteristics of the theory are elaborated in great detail. The essential requirement placed by the theory on all linguistic analyses and interpretation is that it proceed in a straight line from level to level and that no level be jumped over. In the case of machine translation, this means that the process should go from the lowest stratum of the source language through its higher strata to the highest stratum; then the translation should be effected on the highest stratum into the target language; and then the process should continue on downward through the strata of the target language until the lowest stratum is reached for the output. The requirement of going down and up all the strata without regard to possible shortcuts seems to encumber the translation process unnecessarily.

DESCRIPTIVE GRAMMARS

These are approaches to grammar which can be characterized as being weakly model-oriented. This means that, while they are based on a definite theoretical framework, the major emphasis is placed not on that framework, but on the manner of implementing it in the description of a particular language. This emphasis on implementation in the description has led to the conception that the aim of these approaches to linguistics is the development of a discovery procedure—that is, a set of rules by which the conceptual framework can be applied to a given corpus of linguistic data in order to describe these data.

Proponents of formal grammar often reject the aim of developing a discovery procedure as unrealistic. This is correct only if the term "procedure" is taken in a strict logico-mathematical way—that is, if it is assumed that the aim is to develop a completely automatic algorithm for linguistic discovery. While this naive aim may have been professed by some proponents of the descriptive approach to grammar, most descriptive approaches are more modest: they merely stress the need for a greater formalization of the procedures now followed by linguists in attempting to describe a language. Proponents of a descriptive approach usually do not have the illusion that a complete discovery algorithm can be devised, but are not willing to go to the contrary extreme of rejecting the formulation of more explicit investigative procedures as an important aim in linguistics.

In the development of American linguistics, the descriptive tradition has been strongly influenced by the rejection of the use of linguistic meaning as a criterion in the analysis. This, in turn, is based on the very strong rejection of "mentalism" by the founder of the American descriptive tradition, Leonard Bloomfield, under the influence of behaviorism in psychology.

Word-Paradigm Grammars. This approach to grammar is among the oldest in the history of linguistics. Its primary aim is the description of word morphology by means of classifying words into classes and subclasses on the basis of the paradigms that these words may take. A *paradigm* in this connection is a set of affixes (most commonly suffixes) characterizing a given class of words with respect to a particular grammatical category. For instance, a noun paradigm in Russian will be the set of suffixes characterizing that Russian noun with respect to case and number.

It is clear that for purposes of machine translation, the description of word morphology in terms of different paradigms is very important for the development of the grammar codes that are used in the syntax routines. In the case of Russian, word paradigms were found necessary but not sufficient, since other important syntactic properties of words had to be taken into account for inclusion in the grammar code.

Item-and-Process Grammar. This approach to grammar is also largely concerned with morphology, although it could be applied to syntax as well. Instead of describing words by giving their paradigms, in this approach words are described as undergoing a certain grammatical process in order to express grammatical categories. So for instance, it is said that Russian words undergo the process of suffixation in order to express case and number; English strong verbs undergo the process of stem change in order to express tense—and so forth.

It is clear that this approach may be of use particularly for those cases to which a word-paradigm approach does not lend itself because of the absence of a clear-cut paradigm. In machine translation terms, however. it is not very efficient to program a grammatical process into the recognition routine. Thus, the applicability of an item-and-process approach to machine translation seems to be rather limited.

Item-and-Arrangement Grammar. The basic characteristic of this descriptive framework is that it allows only two parameters of linguistic description: items and their arrangement. Thus, a grammatical process is to be represented, not as a process, but as a rearrangement or replacement of items. For instance, English strong verbs are no longer said to express tense by a process of stem change. The item-and-arrangement approach requires that strong verbs be represented as having several different forms occurring in place of each other in order to mark tense distinctions. This approach to grammar seems to be rather restrictive since it is quite clear that not all grammatical phenomena can be described by using only items and their arrangement as the criteria for the description. This restrictiveness is due to the influence of anti-mentalism on American linguistics. The assumption was that a "process" type description was unscientific because it was anthropomorphic. (Only people can "take" things; it is therefore unscientific to say that a verb "takes" a suffix. The correct way of saying it is that the verb stem is "followed" by a suffix.)

An item-and-arrangement grammar thus is clearly inadequate as a tool for describing a natural language. Nevertheless, the nature of computing equipment forces an item-and-arrangement approach upon a machine translation treatment of morphology. The routines for separating suffixes from word stems used in machine translation are an example of item-and-arrangement handling of data, as is dictionary look-up or, for that matter, word order rearrangement. We interpret this to mean that there is no one-to-one correspondence between a linguistic description and a computer treatment of linguistic data.

Immediate-Constituent Grammars. This approach to linguistic description is based on the assumption that linguistic analysis can proceed from the largest linguistic unit, the sentence, down to its smallest ultimate constituents in an orderly progression. The order of the analysis is to proceed at each step only to the immediate constituents of the unit under investigation. At the next step, the results of the previous step are again subjected to an analysis into their immediate constituents and so on until the ultimate constituents are reached. The immediate-constituent approach as applied to syntax is roughly comparable to a phrase-structure approach. Instead of rewriting symbols, however, the immediate-constituent approach is couched in terms of the analysis of a given unit into its constituents. Thus, the immediate constituents of a sentence are its functional components—namely, the subject and predicate. The

immediate constituents of these, in turn, are further components such as modifier and head for the subject, and verbal construction and object for the predicate. The composition of the sentence is ascertained by impressionistic inspection and the sequence of steps of the immediate-constituent analysis is conducted accordingly.

While the conception of a progressive analysis is unquestionably sound, the necessary recognition criteria for the components that are to be singled out at each step are not provided. Thus, the approach lacks the explicitness needed for application to a machine translation program.

Tagmemic Grammar. This is a further development of the American descriptivist conception of levels and of the technique of immediate-constituent analysis. Its basic concept is the tagmeme which is defined as a "slot-filler correlative." In simple terms, the structure of a sentence is represented as a series of functional slots in which may occur appropriate fillers. An example is the subject slot in which appropriate fillers may be nominal phrases or pronouns. The analysis then consists in ascertaining how many of these slots are given in a particular language, and what are appropriate fillers. Recently, an attempt has been made to formalize the tagmemic approach by working out so-called tagmemic formulas. These are intended to represent both the functional slot and the pertinent fillers.

The most attractive feature of this approach is the emphasis on function, which constitutes one of the important conditions to be ascertained in syntactic recognition. The differentiation of slot and filler implies a difference between a linguistic function and its carrier, and thus contributes to the explicitness of linguistic description. On the other hand, here, as in the previously mentioned descriptive approaches, the very important question of recognition signals is ignored.

Glossematic Grammar. This appproach to linguistic analysis was developed in the European tradition and is therefore not influenced by anti-mentalism or behaviorism.

An essential feature of glossematics is the differentiation between language as a system on the one hand and the manifestations of language in actual speech on the other. This is a further development of the distinction between language and speech made by the Swiss linguist, F. de Saussure. The assertion is that the system of a language can be revealed by examining a text (that is, a sample of speech) that represents it. The examination consists in a consecutive series of divisions of the text into smaller and smaller components. The most significant characteristic of a linguistic system is that it is composed of the units of language in a variety of different relations of dependence with each other. One such relation is that of unilateral dependence—that is, unit A presupposes unit B, but not conversely. A grossly oversimplified example of this relation is that of the English article to the noun: the article presupposes a noun for its occurrence, but not conversely; hence, the article is unilaterally dependent on the noun. Another relation is that of mutual dependence in which A presupposes B and B presupposes A. An example of this would be the relation between stem and ending in many Russian words, where neither may occur without the other. Several different types of relations have been worked out in glossematic theory. In addition, a very detailed hypothesis of the different aspects of language and the relations of language to metalinguistic phenomena has been worked out. While many of the details of glossematic theory are much too abstract to be applicable to any practical task, the emphasis on different types of relations is of interest because it corresponds to the emphasis on function mentioned earlier. As in the previously discussed grammars, so in glossematic grammar, no adequate attention is paid to the problem of recognition signals.

Fulcrum Approach to Syntax. This is the approach followed by the author and his associates in the design of syntactic recognition routines for machine translation. It is an attempt to combine the best features of all the descriptive approaches.

A fulcrum grammar is in some ways comparable to an immediate-constituent grammar, but with some significant differences. Where the immediate-constituent approach takes the maximum unit—the sentence—as its point of departure and considers its step-by-step breakdown into components of an increasingly lower order of complexity, the fulcrum approach starts out with the minimum unit—the morpheme (minimum unit of grammatical form) in straight linguistic analysis, the typographical word in language data processing—and considers its gradual fusion into units of increasingly higher orders of complexity, called fused units. A sentence is thus visualized, not as a simple succession of linear components, but as a compound chain of fused units of different orders of complexity variously encapsulated in each other. Syntactic analysis, including the automatic analysis which a machine translation syntax routine must perform, then has as its objective the identification of this encapsulation of fused units by ascertaining their boundaries and functions.

The objective stated above encompasses two tasks in the case of machine translation: one is the design of subroutines for appropriate identification of recognition signals, the second is the design of an overall program in which the recognition of individual elements within the sentence is accomplished in the proper order. The latter is the well-known problem of "traffic rules" as formulated by the proponents of formal grammars.

In the fulcrum approach as applied to Russian, the recognition criteria for the boundaries and functions of particular sentence components consist in the appropriate collocations of the functional potentials of Russian words as recorded in their grammar codes. For instance,

the subject function is recognized on the basis of the following criteria: (1) by the membership of a word in the part of speech of nouns which constitutes potential for functioning as a subject or object; (2) the nominative case form of the noun; (3) the presence in the same sentence of an appropriate verbal expression.

A recognition routine based on the fulcrum approach directs the primary syntactic searches upon those pivot words (fulcra) within the sentence around which other words are centered. The fulcra are those words which contain the maximum amount of grammatical information, and which thereby allow an optimization of further searches within the sentence. Not only is the sentence as a whole assumed to have a fulcrum, but the various constituents of a sentence in turn are assumed to have their fulcra.

The recognition criteria are identified in the order prescribed by the concept of the fulcrum: the recognition criteria for the central element of a structure are attended to first, and only then are the subsidiary elements of each structure searched for. In view of the great variety of structures present in a normal sentence, these searches are conducted not all at once but in an order prescribed by the pass method. This order is established on the basis of analytical knowledge of Russian grammar. The earlier passes in the syntax program serve to resolve ambiguities and to prepare as nearly unambiguous a set of grammar codes as possible. In the subsequent passes, attention can then be turned to the fulcra of the major sentence components.

• Some Remarks on Linguistics and Content Processing

In this author's view, content processing shares with machine translation the need for a linguistic recognition routine which will yield an automatic analysis of the structure of the sentences of the document. In technical terms, this requires the inclusion in the system of a machine dictionary equipped with a grammar code capable of calling appropriate subroutines for syntactic recognition. The machine translation dictionary will have to be bilingual, and the program will have to include provisions for multiple-meaning resolution. For content processing, the dictionary will be monolingual and will have to include, in addition to a grammar code, a semantic code capable of calling appropriate subroutines for content comparison and evaluation. This is because in content processing the recognition of the content of a document is only the first step. The principal objective here is to evaluate the content in order to process it further for a given purpose, such as indexing for storage and later retrieval, or abstracting.

This evaluation requires the automatic inclusion of some kind of relevance criterion by means of which certain portions of the document can be highlighted and other portions can be ignored. The criterion for such an evaluation in the case of information retrieval seems to be the comparability of each particular document to the documents of a related set; common features and differences can serve as a basis for an index in terms of which information can be retrieved in response to a request. In automatic abstracting, various portions of the document are compared and, on the basis of their relative significance, are retained or omitted from the condensed version.[1]

The development of these relevance criteria presupposes a detailed empirical and theoretical knowledge of the semantic system of natural languages. Linguistic research in this area, as was suggested in the Introduction, is still in its early stages, and results of interest to content processing are now first beginning to come in.

References

1. Paul L. Garvin, ed., "A Linguist's View of Language Data Processing," *Natural Language and the Computer* (New York: McGraw-Hill, 1963) p. 110.
2. Cf. Noam Chomsky, *Syntactic Structures* (The Hague: Mouton and Company, 1957) p. 85.

Comparative Indexing: Terms Supplied by Biomedical Authors and by Document Titles*

CLAIRE K. SCHULTZ, WALLACE L. SCHULTZ,
and RICHARD H. ORR

Institute for Advancement of Medical Communication
Philadelphia, Pennsylvania

The original aim of this study was to obtain objective data bearing on the much argued question of whether author indexing is "good." Author indexing of 285 documents reporting biomedical research was scored by comparing the author-supplied terms (author set) for each paper with a criterion set of terms that was established by asking a group of 12 potential users to describe the same document. Terms in the document title (title set) were scored similarly. The average author set contained almost half of all the terms employed by more than one member of the user group and scored 73% of the maximal possible score, as compared with 44% for the average title set. When judged by the method and criterion employed here, author indexing is substantially better than indexing derived from document titles. The findings suggest that indicia supplied by an author should serve scientists in biomedical disciplines other than his own about as well as they serve his disciplinary colleagues. The general method developed for measuring indexing quality may represent a practical yardstick of wide applicability.

Introduction

ORIGINAL AIM OF STUDY

In 1958, Lee stated, "Whether or not authors can supply indicia terms sufficiently good that professional indexers can quickly transform them into standard entries is questionable, but the experiment should be made."[1] The question he posed has two parts—the first relating to the quality of author indexing, and the second to the ease with which author-supplied indexing terms can be converted into some controlled indexing language. The "experiment" he urged—having authors supply indexing terms for their papers—is now under way in several broad fields, for example, in the biomedical sciences [2] and in physics [3]; and wherever journal editors have established the policy, authors have soon become accustomed to submitting, along with their manuscript, a list of terms they consider appropriate for indexing purposes. The experience to date has answered pragmatically the second part of Lee's question in that author-supplied indicia are being standardized by human editing[3] and even by computer[4] as a practical, routine operation. The first and more important part of his question, however, is still unanswered. Strong opinions on the quality of author indexing are common,† but the question has received little systematic study.

In an attempt to collect objective data that might help resolve the quality question, at least for author indexing of biomedical research reports, we began the present study in early 1962. At that time, the use of author-submitted terms in compiling an index for a large and important group of documents reporting on-going biomedical research had become routine[0], and a unique opportunity presented itself: we had a chance to enlist as highly motivated participants, each willing to devote many hours to the study, a number of the prime users of these documents—biomedical scientists actively engaged in research. Given a corpus of documents with author indicia, and the prospect of intensive cooperation from such a user group, the first problem was

* This work was supported, in part, by Public Health Service Grant GM-09166 from the National Institute of General Medical Sciences.

† The following quotations from letters-to-the-editor in *Science* illustrate the polar views on this question: "Since the greatest authority on any item of literature is the author, is he not the one best able to classify the item properly?"[5] and ". . . authors are seldom qualified to do indexing. Only occasionally can they do half as well as an experienced professional indexer."[6]

how to use these resources to determine whether the author indexing was "good" or "bad."

CRITERIA FOR INDEXING

The most direct test of the quality of any subject indexing is, of course, how well it serves those who seek answers to questions, or documents relevant to their interests. Cleverdon [7], and others [8], are developing ways to assess indexing by measuring its effectiveness and efficiency for retrieval. However, since indexing is only one in a sequence of operations that must be performed in any complete system for storing and retrieving documents, its quality cannot be measured in terms of retrieval performance unless the major variables affecting other operations are controlled. Developing a "laboratory" model of a storage and retrieval system, such as is required to test indexing under controlled conditions, has proved to be a major undertaking; and to date, at least, it has been necessary to incorporate in the testing model certain artificialities that facilitate control of variables but, at the same time, decrease confidence that the test results accurately reflect the performance of real systems.

Because of the theoretical and practical difficulties in testing indexing by retrieval performance, another type of criterion has been adopted in most efforts to measure the quality of indexing; one set of indexing terms for a document has been considered a standard against which "unknowns" (i.e., other sets of indicia for the same document) are compared. The standard or criterion set of terms has been established by some mode of indexing that is generally considered to produce a useful, if not optimal, product. Borko, for example, evaluated computer-selected terms by comparing them with the terms selected for the same document by two subject-matter specialists; [9] and Stevens compared machine indexing with that performed by a professional indexer [10] The validity of this type of criterion rests on a basic assumption, usually implicit, that, because a mode of indexing has been, and continues to be, widely employed by information services (e.g., assignment of terms by professional indexers), it must provide users with an "acceptable" level of retrieval performance.

Although criteria validated by acceptability may be adequate for certain purposes, they have serious shortcomings for general use in quantitative comparisons of indexing quality. First, users of information services frequently have little freedom to choose among alternatives and must accept whatever service is available or do without; and what constitutes acceptable indexing varies from one field of science to another. Second, the direct measurements of retrieval performance made to date [7] suggest how great and variable the difference between acceptable and optimal indexing may be.

BASIC PREMISE OF PRESENT APPROACH

We also adopted a criterion other than observed performance in retrieval, but one whose validity was not based on the truism that a widely used mode of indexing gives an acceptable product. We needed a standard that could be considered to represent optimal indexing, and against which author-supplied indicia, or terms produced by any other mode of indexing, could be assessed without prejudice. A reasonable premise for developing such a standard seemed to be the rule-of-thumb "law" that, other things being equal, retrieval performance increases as indexing corresponds more closely with the terminologies and viewpoints of those who use the indexing. We reasoned that, if we could find a practical way to establish a criterion set of indexing terms for a document by determining how the users themselves would describe the document, this set might constitute an appropriate standard for measuring the probable utility, to these users, of other sets of indexing terms for the same document. Ideally, the group employed to establish criterion sets should be a representative sample of the scientists most likely to use the indexing to be assessed; and each member of this group should characterize the documents in whatever ways, and by whatever terms, were natural for him.

DESIGN OF INITIAL PHASE OF STUDY

The basic design of the first investigation that evolved from this approach may be summarized as follows: author indexing for a sample of documents covering a wide range of current biomedical research was evaluated by the extent to which the author-supplied terms corresponded with terms used by a group of scientists actively engaged in biomedical research, each of whom independently described the same documents.

HYPOTHESES

The main hypothesis to be tested was that, as rated by the criterion we adopted, author indexing is of "good" quality. This paper presents the findings bearing on the main hypothesis and on two additional hypotheses regarding author indexing: (1) that indexing terms supplied by an author serve scientists in his own discipline better than scientists in other disciplines, and (2) that author-supplied indicia represent better indexing than terms derived from document titles.

LATER PHASES OF COMPARATIVE INDEXING STUDY

In the course of testing these hypotheses, we became interested in the possibility that the approach adopted, and the general method developed for evaluating indexing, might have wider applicability. This interest led us to explore the variables of the method itself. In addition, the data we had collected seemed to show some interesting patterns in the way scientist-users describe documents; and this observation stimulated further work. These two outgrowths of the original investigations will be reported later, as will another phase in our continuing study of comparative indexing—an analysis of how experienced, professional indexers handle the same documents used in the present assessment of author indexing,

and an attempt to identify the differences and similarities between "amateur" indexers (authors and scientists engaged primarily in research) and professional indexers.

Methods, Material, and Subjects

ORGANIZATIONAL SETTING OF STUDY

The Federation of American Societies for Experimental Biology provided the organizational setting for the author-user population; and since certain features of this setting affected the planning and execution of the study, a brief description follows:

Each year about 3,000 10-minute oral reports of current biomedical research are given at the annual meeting of the Federation. This national convention is the largest meeting of biomedical scientists, and the work presented is an excellent cross section of U.S. biomedical research. The Federation consists of six societies, each representing a major biomedical discipline—biochemistry, immunology, nutrition, pathology, pharmacology, and physiology. Only members of these societies may present unsolicited papers at the Federation meeting; and the speaker must submit to his society a short summary (250 words or less) of what he intends to say.* These summaries are published in the March-April issue of *Federation Proceedings,* which is distributed before the annual convention.

Although the document submitted is commonly referred to as an "abstract," the term is a misnomer in that this type of summary is not usually produced by abstracting a pre-existing document. Authors most commonly prepare the summary *before* they have written out the full text of their oral presentation. When published, these "anticipatory abstracts," therefore, represent *primary* documents—condensed, preliminary reports that may or may not be followed at some later time by publication of a more detailed report.[11]

Since 1960, the author of a paper to be given at this meeting has been required to fill out an author-indexing form and to submit this form along with his summary. The Federation uses the data thus collected to facilitate the grouping of similar papers for scheduling meeting sessions, to classify the summaries into sections for printing in the annual meeting issue of *Federation Proceedings,* and to compile a subject index to the summaries.†

SAMPLE OF AUTHOR INDEXING

Our sample of author indexing was drawn from the 2,854 author-indexing forms submitted for the 1962 meeting of the Federation; Fig. 1 reproduces part of the form used for this particular meeting. The sample comprised the completed author-indexing forms for every tenth document published in the 1962 meeting issue of *Federation Proceedings* (Vol. 21, No. 2, March-April 1962). In this systematic sample, each of the six societies is represented 9 to 11% of all the forms submitted to it; of the 285 forms in the sample, 90 had been submitted to the American Society of Biological Chemists, 81 to the American Physiological Society, 35 to the American Society for Pharmacology and Experimental Therapeutics, 31 to the American Society for Experimental Pathology, 26 to the American Association of Immunologists, and 22 to the American Institute of Nutrition. Because of the rules governing eligibility and submission, this distribution reflects the disciplines of the authors who contributed to the sample.

CRITERION GROUP

The group of users employed to establish the standard for indexing, which we will call the "criterion group," was a selected sample of the authors' peers. Two active research workers from each of the six disciplines represented in the Federation were selected by Dr. Milton Lee, Executive Officer of the Federation, on the basis of their recognized standing in the research community and of his assessment of the likelihood that they would be willing to participate in the study. To facilitate holding a meeting at which uniform instructions and uniform answers to common questions could be given, the original selection was limited to scientists in the Bethesda area (the National Institutes of Health and the Naval Medical Research Institute). One of the 12 individuals originally selected had to withdraw; he was replaced by a scientist who worked in a pharmaceutical company.

Each member of the criterion group was asked to characterize each of the 285 documents corresponding to the completed forms comprising the sample of author indexing and to record his response for each document on a fresh copy of the author-indexing form. The criterion group was instructed to use as many terms as they felt necessary to characterize a document. The following points were emphasized: (1) that the responses should reflect their *personal* viewpoints and terminology, (2) that they should follow the classification and terminology suggested on the printed form *only* to the extent they found it natural and useful, and (3) that the results would be used in considering how the classification and terminology suggested on the printed form might be *revised* to accommodate better the different viewpoints and usage of the six disciplines in the Federation. From the Federation's standpoint, the latter use of the data was the primary reason for enlisting the group's cooperation; our intention to use the data also to evaluate author indexing was not expressed to the criterion group. Although the oral instructions stressed freedom to use

* The rules on eligibility and submission are considerably more complex than this statement would indicate, and vary somewhat from one society to another. In some cases, a nonmember, if he is "sponsored" by a member, may present a paper; in others, the speaker may be a nonmember who is a coauthor of a paper reporting a member's work.

† This description of the uses of the author-indexing form holds for the 1962 meeting. Since then, changes have been made in how the summaries are ordered for printing; and for the 1963 and 1964 meeting issues, in addition to author-supplied indicia, substantive words and phrases from document titles were used in compiling the index.

Please study the subject-category list before marking. The list will be used primarily for the arrangement of the abstracts and for the production of the subject index to the abstracts. A secondary use will be for aid in programming.

Place the number "1" in the box at the left of the most specific category which classifies the area of your paper; the number "2" in the box at the left of the next most specific category. Do not mark more than two categories.

In the blanks at the end of the subject-category list, please supply four or more additional descriptive terms (words or short phrases) which can be used, besides the subject categories already selected, for further classifying and indexing the content of your paper. The terms you supply should preferably be nouns. Generic names of chemical compounds and drugs should be used, rather than trade names or jargon.

SUBJECT CATEGORIES

☐ 001 **Amino Acids**
☐ 002 Metabolism
☐ 003 Nutrition
☐ 004 Synthesis
☐ 005 **Antigen-Antibody Reactions**
☐ 006 Cross Reactions
☐ 007 Haptens
☐ 008 Immunofluorescence
☐ 009 In Vivo Reactions
☐ 010 *Cellular*
☐ 011 *Pathogenetic*
☐ 012 Non-specific Factors
☐ 013 *Complement*
☐ 014 *Properdin*
☐ 015 Precipitation
☐ 016 *Diffusion*
☐ 017 *Immunoelectrophoresis*
☐ 018 *Quantitation*
☐ 019 **Antigens; Antibodies**
☐ 020 Antibody Formation
☐ 021 Determinants
☐ 022 Microorganisms
☐ 023 *Bacteria*
☐ 024 *Rickettsia*
☐ 025 Polysaccharides
☐ 026 Proteins
☐ 027 Toxins
☐ 028 Transplantation
☐ 029 *Autoantibodies*
☐ 030 *Tissue antibodies*
☐ 031 **Biological Oxidations**
☐ 032 Cytochromes
☐ 033 Electron Transport

☐ 040 Coagulation
☐ 041 *Agents; factors*
☐ 042 *Fibrinolysis*
☐ 043 *Platelets*
☐ 044 Erythrocytes
☐ 045 *Destruction*
☐ 046 *Metabolism*
☐ 047 Groups
☐ 048 Hematopoiesis
☐ 049 Hemoglobin
☐ 050 Leukocytes
☐ 051 *Leukemia*
☐ 052 Plasma Proteins
☐ 053 *Albumin*
☐ 054 *Globulins*
☐ 055 Storage
☐ 056 **Body Water**
☐ 057 **Bone**
☐ 058 **Carbohydrates**
☐ 059 Chemistry
☐ 060 Metabolism
☐ 061 *Citric acid cycle*
☐ 062 *Glycolysis*
☐ 063 *Hexose phosphate path*
☐ 064 *Monosaccharide conversions*
☐ 065 *Polysaccharides*
☐ 066 *Small cycles*
☐ 067 Photosynthesis
☐ 068 **Cardiovascular System**
☐ 069 Atherosclerosis
☐ 070 *Experimental*
☐ 071 *Nutritional*
☐ 072 *Pathophysiology*
☐ 073 Blood Flow
☐ 074 *Cerebral*
☐ 075 *Coronary*

☐ 084 *Shock*
☐ 085 Blood Vessels
☐ 086 *Capillary exchange*
☐ 087 *Venous return*
☐ 088 *Wave transmission*
☐ 089 Blood Volume
☐ 090 *Hemorrhage*
☐ 091 *Transfusion*
☐ 092 Cardiac Drugs
☐ 093 Cardiac Muscle
☐ 094 *Disorders*
☐ 095 *Electrocardiography*
☐ 096 Cardiac Output
☐ 097 *Control*
☐ 098 *Measurement*
☐ 099 CV Disease
☐ 100 Edema
☐ 101 Lymph
☐ 102 **Cell Structure; Function**
☐ 103 Active Transport
☐ 104 Cell Membranes
☐ 105 Cytoplasm
☐ 106 Microsomes
☐ 107 Mitochondria
☐ 108 Nuclei
☐ 109 **Cell, Tissue Culture**
☐ 110 Cell Antigens
☐ 111 Metabolism
☐ 112 Neoplasms
☐ 113 Nucleic Acids
☐ 114 **Chemotherapy**
☐ 115 Bacterial
☐ 116 Cancer
☐ 117 Parasitologic
☐ 118 **Connective Tissue**
☐ 119 Disorders

☐ 127 Site
☐ 128 **Drug Metabolism**
☐ 129 **Endocrines**
☐ 130 Adrenal Cortex
☐ 131 Adrenal Medulla
☐ 132 Anterior Pituitary
☐ 133 *ACTH*
☐ 134 *Control of secretion*
☐ 135 *Gonadotropin*
☐ 136 *Somatotropin*
☐ 137 *TSH*
☐ 138 Brain Hormones
☐ 139 Glucagon
☐ 140 Insulin
☐ 141 *Diabetes mellitus*
☐ 142 *Mode of action*
☐ 143 Parathyroid
☐ 144 Posterior Pituitary
☐ 145 *Diabetes insipidus*
☐ 146 Sex Hormones
☐ 147 *Androgens*
☐ 148 *Estrogens*
☐ 149 *Progestogens*
☐ 150 Thyroid
☐ 151 *Iodine metabolism*
☐ 152 *Regulation*
☐ 153 *Thyroxine*
☐ 154 **Energy Metabolism**
☐ 155 **Environment**
☐ 156 Adaptation
☐ 157 Air Pollution
☐ 158 Altitude
☐ 159 Hibernation
☐ 160 Hyperthermia; Heat

☐ 210 *Synthesis*
☐ 211 *Transport*
☐ 212 Phospholipids
☐ 213 *Metabolism*
☐ 214 *Synthesis*
☐ 215 Sterols
☐ 216 *Metabolism*
☐ 217 *Synthesis*

☐ 262 Neurochemistry
☐ 263 Pain
☐ 264 Peripheral Nerves
☐ 265 Reflexes
☐ 266 *Axon*
☐ 267 *Conditioned*
☐ 268 Spinal Cord
☐ 269 **Nitrogen Metabolism**

☐ 320 Control
☐ 321 Disorders
☐ 322 Diuresis; Diuretics
☐ 323 Electrolyte Excretion
☐ 324 Glomerular Filtration

☐ 366 *Hepatitis*
☐ 367 **Vitamins**
☐ 368 B
☐ 369 B_{12}
☐ 370 C
☐ 371 Fat-Soluble
☐ 372 Folic Acid
☐ 373 Unidentified

ADDITIONAL DESCRIPTIVE TERMS

....
....
....

FIG. 1. "Author-Indexing" Form. The form consisted of 2 pages; here the lower part of the first page and the top of the second page have been omitted.

any term they wished and any number of terms, the criterion group was not explicitly told to ignore the directions at the top of the author-indexing form (Fig. 1), which called for selecting two, and only two, of the numbered categories, and for writing in four or more "additional descriptive terms."

Most of the work of the criterion group was completed during the summer of 1962. One member was unable to finish the entire assignment and responded to only 91 of the 285 documents. Several others failed to respond to a few of the documents, either inadvertently or because they felt that a particular document was too far outside their area of special competence for them to be confident of their ability to characterize it meaningful.* However, of the 3,420 completed forms that would have resulted had all 12 members of the criterion group responded to each of the 285 documents, 3,194 (93%) were returned; these forms constituted the raw data for establishing the criterion sets of indexing terms.

FLAW IN EXECUTING THE STUDY DESIGN

The design of the study called for each member of the criterion group to respond independently (i.e., without knowledge of how others in the group responded or of how the author had indexed his paper). After the original data had been analyzed, while preparing to report the findings, we realized that the latter condition had been comprised by an oversight. Rather than giving members of the criterion group 285 separate documents, we had supplied each with a copy of the 1962 meeting issue of *Federation Proceedings* and indicated by red brackets which of the 2,854 documents therein comprised the sample to be characterized. In *Federation Proceedings,* the summaries of papers given at the meeting had been classified into sections according to whichever of the 373 subject categories on the author-indexing from the author had checked as his first choice (i.e., as the category he felt classified his paper most specifically). Terms designating the subject categories served as headings for the sections in *Federation Proceedings.* The "section heading" terms, therefore, represented author-supplied indicia; and to the extent that these terms had been noticed by the criterion group, they could have influenced the group's use of terms and hence the resulting criterion sets.

Even if the members of the criterion group had been interviewed immediately after completing their assignments, it would have been difficult to determine how much they had actually been "led" by the section heading terms, which were also carried as running heads on the pages of the given section. In view of the considerable time that had elapsed before the oversight was recognized, this technique would obviously not resolve the question definitively. To satisfy our curiosity, however, we sent the 12 scientists a simple questionnaire that reviewed how the documents had been supplied and asked for a yes-or-no answer to the question: "Were you aware of the content of the running heads of the pages as you worked?" Ten questionnaires were returned; seven answered "Yes," (one volunteered the comment that he "made a conscious effort to avoid the running heads"), and three "No."

DATA TREATMENT

Recording Term Uses. As can be seen in Fig. 1, the arrangement of subject categories on the author-indexing from was, in part, hierarchical; as many as three levels of specificity are displayed. For example, subject category No. 058 is displayed as generic to categories 059 through 067, and category 060 is generic to categories 061 through 066. When a subject category is removed from the context of this display, it must be supplemented by adding the terms for any more generic categories in the given hierarchy, e.g., the "expanded" term for category 061 is "Carbohydrates—Metabolism—Citric acid cycle." We recorded, in such an expanded form, any terms that authors or members of the criterion group had chosen from the printed list of subject categories.

Free-language write-ins at the end of the form were edited before further processing. Both authors and criterion group often used phrases containing prepositions, conjunctions, verbs, and articles, as well as nouns and adjectives; only substantive words were considered as indexing terms. After reviewing all the terminology that had been used by authors and criterion group to describe a given document, C. K. Schultz, who had been studying the linguistic patterns of Federation scienists for several years, made the decisions on standardization. The principal rules were to standardize when the same word was used in singular and plural forms or when two phrases or words appeared to be equivalent *as used by Federation scientists* (e.g., the expressions, "maple syrup urine" and "maple syrup urine disease," were standardized, as were "hepatic disorders" and "liver disorders"). In some cases, a write-in phrase was found to be equivalent to one of the terms designating a subject category in the printed list. The aim of the limited standardization performed was to preserve all non-trivial variation.†

The titles of documents were handled in the same way

* Another departure from ideal execution of the study design occurred when three members of the criterion group, who could not otherwise complete the assignment (which probably required a minimum of 20 to 80 hours), enlisted the help of one or more of their co-workers. For the present purposes, we have assumed that, in each case, these co-workers were in the same discipline as the given criterion group members and all the forms he returned have been treated as if they were his responses.

† Investigators who wish to analyze the unedited terminology of the write-ins and document titles, or to obtain a complete set of the 285 term-use matrices, which present all the edited data, should contact the authors.

Table 1. Term-Use Matrix for Document No. 100

Term[a]	Biochemist A	Biochemist B	Physiologist A	Physiologist B	Pharmacologist A	Pharmacologist B[b]	Pathologist A	Pathologist B	Immunologist A	Immunologist B	Nutritionist A	Nutritionist B	Term-Weighting Factor	Author	Title
Chemotherapy—Cancer (116)		x	x		x		x	x		x	x			x	
Neoplasms				x	x			x				x	16	x	x
Mast cells				x					x	x			9	x	x
Cell, tissue culture—Neoplasms (112)	x				x								4	x	
Chemotherapy (114)							x				x		4		
Benadryl				x						x			4		
Histamine-antihistaminics (195)				x					x				4	x	
Bacteria				x									1		
Cell structure; function (102)	x												1		
Cell, tissue culture (109)	x												1	x	
Drug action (123)										x			1		
Hosts								x					1		
Metabolism				x									1	x	
Peritoneal fluid													0		x
Mice													0		x

(Column group heading over the Biochemist to Nutritionist columns: Criterion Group.)

[a] The parenthetical numbers following certain terms refer to the code numbers of the corresponding subject categories listed on the author-indexing form used by both authors and criterion groups; in these cases, the term given here may represent an "expansion" of the term that appeared on the form (see text). Terms not followed by a number are "write-ins," i.e., words or phrases not included among the subject categories listed on the printed form.
[b] This member of the criterion group did not respond to document No. 100.
[c] This term served as the heading for the section in *Federation Proceedings* in which document No. 100 appeared.

Table 2. Characteristics of Criterion Sets

Authors' Discipline	No. Documents (1)	All Terms: Mean No./Set (2)	All Terms: Mean Weight (3)	All Terms: % with Weight of 1 (4)	SH Terms: Mean No./Set (5)	SH Terms: Mean Weight (6)	Terms other than SH Terms: Mean No./Set (7)	Terms other than SH Terms: Mean Weight (8)
Biochemistry	90	12	10	49%	1	52	11	6
Physiology	81	11	11	45	1	57	10	7
Pharmacology	35	12	10	44	1	52	11	6
Pathology	31	10	12	45	1	58	9	8
Immunology	26	10	11	47	1	60	9	6
Nutrition	22	14	10	49	1	53	13	6
All Documents	285	11	11	47	1	55	10	7

All figures are rounded off to the nearest whole number.
"All Terms" in given sets) = "SH Terms" (terms serving as section headings) + "Terms other than SH Terms"; in columns (2), (5), and (7), "Mean No./Set" = Total No. of terms (of the type specified in the heading over the column) in the given sets ÷ No. of sets, i.e., by the No. of documents shown in column (1); in columns (3), (6), and (8), "Mean Weight" = sum of weighting factors for terms (of the types specified in the heading above the column) in the given sets ÷ No. of such terms in the given sets; in column (4), "% with Weight of 1" = No. of terms with 1 as a weighting factor × 100 ÷ No. of all the terms in the given sets.

as the write-ins by authors or by criterion-group members.

A term-use matrix, similar to that shown in Table 1, was created for each document (see footnote to the paragraph above). Each term use by each of the 12 members of the criterion group and by the author was indicated by an X; the "presence" of a term in the title was similarly recorded. The two members of each of the six disciplinary pairs making up the criterion group were designated *A* and *B*. In the example shown, 7 of the 15 terms listed represent subject categories displayed on the printed author-indexing form.

Weighting of Terms. The criterion set of terms for a document consists of all the *different* terms used by members of the criterion group to describe that document; thus 285 criterion sets were established, one for each document. For document No. 100, the criterion set contained 13 terms. Each term in a criterion set was assigned a weighting factor equal to the square of the number of criterion group members who had used it to describe the given document. The weighting procedure may be illustrated with the data on document No. 100; for each term, the figure in the column headed "Term-Weighting Factor" was obtained by squaring the total number of X's appearing to the left of this column in the given row. Note that the weighting of a term was *not* affected by whether the author had or had not included it among the indicia he supplied, or by whether it was supplied by the document title. Terms supplied by the author, or by the title, that had not been used by at least one member of the criterion group were *not* included in the criterion set; we will refer to such terms as "zero terms" since they were given a weighting factor of 0.

To test the hypothesis that the indexing terms supplied by an author serve scientists in his own discipline better than scientists in other disciplines, we employed a sub-set of the criterion set for each document. This "disciplinary" sub-set contained *only* the terms used by one or both members of the criterion group who were in the *same* discipline as the author.* Terms in the sub-set were weighted in the same way the full criterion set had been weighted, except that only term uses by the appropriate two members of the criterion group were counted; therefore, the weighting factor for any term in the sub-set was either one or four. To illustrate with the data in Table 1: the disciplinary sub-set for document No. 100, where the author was a pathologist, consists only of the four terms used by pathologists *A* and *B*. For all other analyses, the full criterion sets, established and weighted by the term uses of all 12 members of the criterion group, were employed.

Raw Scores of Author and Title Sets. For each document, the terms supplied by the author constitute another term set ("author set"), and the terms supplied by the title of the document comprise a third set ("title set"). Thus for each of the 285 documents in the sample, there are 3 independently established sets of terms—a criterion set, an author set, and a title set. Each author and title set was scored on the basis of the criterion set for the same document. The scoring procedure was to add, for each author and each title set, the previously established weights for all the terms it shared with the criterion set; this sum was the "raw score" of the given author or title set. The author set for document No. 100 received a raw score of 84 (49+16+9+4+4+1+1), and the title set scored 25 (16+9+0+0).

It is important to note that, in the scoring and weighting procedures, each term was treated as if it were unrelated to any other term. Thus, "basal metabolism" and "metabolism" were considered to be independent terms, and generic-specific relations between terms (e.g., "chemotherapy" and "chemotherapy-cancer") were ignored.

Standardization of Raw Scores. Since the number of terms, and the weighting of these terms, varies from one criterion set to another, the constraints on the raw scores also vary; and a raw score of 50 for an author set, when the corresponding criterion set contains 16 terms with a total weight of 149, is not directly comparable with the same raw score when the criterion set has only 6 terms with a total weight of 86. To facilitate comparisons, we converted the raw scores into percentages of the highest score that *could* be awarded ("maximal score"), i.e., the sum of the weighing factors, for all terms in the criterion set. For document No. 100, if the author set had contained all of the 13 terms in the criterion set, its raw score would have equaled the maximal score for this document (96); the actual raw score (84) was 88% of the maximal score.

• Data

CRITERION SETS

General Characteristics. Of the 285 criterion sets, the smallest contains 4 terms and the largest 27; the weighting factors for individual terms range from 1 to 144; and the sum of the weights for all the terms in a set varies from 40 to 220. Table 2 shows the mean number of terms per set, and the mean weight per term, for each of the six disciplinary groups of documents and for the entire document sample; this analysis is broken down by the two types of terms in the criterion sets—terms that served as the section headings for the documents (SH terms) and terms other than SH terms. For the document sample as a whole, almost half

* As implied in the description of the sample of author indexing, we assumed the author's discipline to be that of the society to which he submitted his paper. The complex rules on submission permitted certain exceptions to the simple relation assumed. Where a document had several authors, all of them may not have been in the same discipline; or where a paper was "sponsored," the author's discipline could have been different from that of the sponsor. Such exceptions were relatively uncommon, however, and should not affect the findings materially.

of all the terms in a criterion set were used by only one member of the criterion group (see "% with weight of 1," Table 2); when such terms are excluded, the mean number of terms per set is 6, rather than 11 as shown in Table 2.

Section Heading Terms. The SH term was specified on 62% of the 3,194 forms completed by the criterion group and, on the average, was used by 7 to 8 members of the group in describing a document.* In 69% of the criterion sets, the SH term was used by more members of the criterion group than any other term. For all criterion sets, the mean weight of the most used term *other than* the SH term is 28; this weight indicates that it was used by 5 to 6 members of the criterion group.

RESPONSES OF INDIVIDUAL MEMBERS OF CRITERION GROUP.

On the average, each member of the criterion group employed 2.5 terms in describing a document, and the sum of the weights for the terms he used averages 67 (mean weight per term, 27). The differences and similarities among the members of the criterion group with regard to how they described the documents will be analyzed in a later paper.

AUTHOR AND TITLE SETS

General Characteristics. The analyses of author and title sets summarized in Table 3 allow comparisons with the general characteristics of the criterion sets; columns (2), (3), (4), (8), (9), (10), and (11) are analogous to columns (2), (3), (4), (5), (6), (7), and (8), respectively, in Table 2. Whereas, all terms in the criterion sets are, by definition, weighted terms, author and title sets can contain zero terms, i.e., terms not found in the corresponding criterion sets. In columns (6) and (7) of Table 3, zero terms are excluded from the calculations.

Scores. Table 4 compares the scores of author and title sets and indicates how much each term contributed to these scores. The median score for author sets is 76%, with roughly one-sixth of the scores falling below 50% and one-fifth above 89%. The figures in columns (5) and (9) express the proportion of all terms in the criterion sets that are also in the author and title sets. For all documents, the author sets include, on the average, 33% of the terms in the corresponding criterion sets; when terms used by only one member of the criterion group are excluded, this proportion increases to 48%; the corresponding figures for title sets are 24% and 35%.

When each author set is scored on the basis of the disciplinary sub-set that contains only terms used by the two members of the criterion group in the same discipline as the author, the mean score is 56%. This is materially lower than the mean score of 73% shown in Table 4, which is based on the full criterion sets and reflects the term uses of all 12 members of the criterion group. When SH terms are not counted, the scores on the two different bases become 37% and 50%, respectively, as compared with 56% and 73%.

* It is of some interest that the criterion group member who, in responding to the questionnaire concerning awareness of the section headings, volunteered the information that he consciously "avoided" them, used the SH term in describing 61% of the documents, as compared to an average of 62% for all members of the criterion group.

• Discussion

ADVANTAGES OF MATERIAL

The Federation meeting was an excellent source of material for a study of this kind. The unique advantages of this material warrant a brief summary. Papers presented at this meeting cover the entire spectrum of biomedical research, and author indexing has become routine. The summaries submitted for this meeting are *primary* documents; but because they are very brief, we could assume that the members of the criterion group read each document in its entirety and could ask each of these scientists to characterize a large number of documents. In view of the breadth and importance of the meeting, almost any active biomedical scientist, in this country at least, could be considered a potential or actual user of the author indexing employed to prepare the subject index for the meeting issue of *Federation Proceedings;* this simplified the selection of a criterion group. The linguistic patterns of Federation members had been studied for several years, and we knew from experience that their free-choice language could be handled. Finally, the author-indexing form developed by the Federation, while encouraging authors and criterion groups to use language that was natural for them, also channeled terminology enough to reduce the amount of standardization needed to facilitate analysis.

WEAKNESSES IN STUDY DESIGN AND EXECUTION

In addition to the flaw in execution described under "Methods, Materials, and Subjects," the present study has other weaknesses. From the beginning, we realized that any criterion group we could marshal with the resources available from the study would not be a representative sample of the total population that uses the indexing to be evaluated; therefore, we settled for a selected sample of scientists who could be considered the authors' peers, if not typical biomedical research workers.

Another weakness is one common to all studies of indexing where terminology is standardized by human editing. Any decision on whether a word is "substantive," whether a variation is "trivial," or whether two words are "equivalent" is inherently subjective; and the task of making completely explicit the rules followed in these types of decisions has defied generations of indexers attempting to train others in the art of index-

TABLE 3. Characteristics of Author and Title Sets

Authors' Discipline	No. Documents (1)	All Terms: Mean No./Set (2)	All Terms: Mean Weight (3)	All Terms: % with Wt. of 1 (4)	Zero Terms: % of All Terms (5)	All Weighted Terms: Mean No./Set (6)	All Weighted Terms: Mean Weight (7)	SH Terms: Mean No./Set (8)	SH Terms: Mean Weight (9)	Weighted Terms Other than SH Terms: Mean No./Set (10)	Weighted Terms Other than SH Terms: Mean Weight (11)
A. Author Sets											
Biochemistry	90	5.0	17	16%	22%	3.9	22	1.0	52	2.9	11
Physiology	81	4.5	20	16	21	3.5	25	1.0	58	2.5	12
Pharmacology	35	5.5	15	18	18	4.5	19	1.0	52	3.5	9
Pathology	31	3.9	24	16	17	3.3	29	1.0	58	2.3	16
Immunology	26	4.3	21	12	22	3.4	27	1.0	60	2.4	13
Nutrition	22	5.0	17	11	28	3.6	24	1.0	53	2.6	13
All Documents	285	4.8	18	16	21	3.7	25	1.0	55	2.7	12
B. *Title Sets*											
Biochemistry	90	3.2	13	16%	20%	2.6	16	0.2	85	2.4	11
Physiology	81	3.3	18	15	17	2.8	22	0.4	80	2.4	13
Pharmacology	35	3.6	13	17	20	2.9	16	0.3	59	2.6	11
Pathology	31	2.8	22	20	9	2.5	24	0.3	76	2.2	16
Immunology	26	3.8	15	14	17	3.1	19	0.4	74	2.8	12
Nutrition	22	4.0	14	15	20	3.2	17	0.3	78	2.9	11
All Documents	285	3.4	16	16	18	2.8	19	0.3	77	2.5	12

All figures are rounded off to the nearest whole number except those in columns (2), (6), (8), and (10), which are to nearest first decimal place.
"All Terms" (in given sets) = "Zero Terms" (terms with a zero weighting factor) + "All Weighted Terms"; "All Weighted Terms" = "SH Terms" (terms serving as section headings) + "Weighted Terms other than SH Terms"; figures for "Mean No./Set," "Mean Weight," and "% with Wt. of 1" are calculated as in Table 2: in column (5), "% of All Terms" = no. of zero terms × 100 ÷ total no. of all terms in the given author or title sets.

TABLE 4. Scores for Author and Title Sets

Authors' Discipline	No. Documents (1)	Author Sets: Mean Raw Score (2)	Author Sets: % Max. Score (3)	Author Sets: % Max. Score/Term (4)	Author Sets: % Max. No. Terms (5)	Title Sets: Mean Raw Score (6)	Title Sets: % Max. Score (7)	Title Sets: % Max. Score/Term (8)	Title Sets: % Max. No. Terms (9)
Biochemistry	90	85	74%	15%	33%	42 (6)[a]	37% (5%)[a]	12% (2%)[a]	22%
Physiology	81	88	72	16	33	60 (12)	49 (10)	15 (3)	26
Pharmacology	35	84	74	13	38	46 (6)	40 (5)	11 (1)	24
Pathology	31	93	73	19	32	60 (8)	47 (6)	17 (2)	25
Immunology	26	92	80	19	32	58 (8)	50 (7)	13 (2)	30
Nutrition	22	86	64	13	26	54 (11)	40 (8)	10 (2)	23
All Documents	285	88	73	15	33	52 (9)	44 (7)	13 (2)	24

All figures are rounded off to the nearest whole number.
[a] Figures in parentheses are for only those terms in the title sets that are *not* also in the corresponding author sets.
In columns (2) and (6), "Mean Raw Score" = sum of the weighting factors for all terms in given author or title sets ÷ No. of sets, i.e., by No. of documents shown in column (1); in columns (3) and (7), "% Max. Score" = sum of the weighting factors for all terms in given author or title sets × 100 ÷ sum of the weighting factors for all terms in the corresponding *criterion* sets; in columns (4) and (8), "% Max. Score/Term" = "% Max. Score" shown in columns (3) and (7), respectively ÷ mean No. all terms in given author or title sets [column (2), Table 3]; in columns (5) and (9), "% Max. No. Terms" = No. of all *weighted* terms in given author or title sets [column (6), Table 3] × 100 ÷ mean No. of all terms in corresponding *criterion* sets [column (2), Table 2].

ing. In studies of indexing, this inability to state rules for standardization completely and unambiguously poses a serious problem; without such rules there is no assurance that a study can be replicated, and the investigator cannot be certain that he has not unconsciously influenced the results. Although the extent of standardization in the present study was relatively limited, we find ourselves in the same position as other investigators in not being able to state, with the requisite clarity and precision, the rules that were followed. For studies of indexing, it would seem that the only way to solve this problem definitively is either to avoid any standardization, or to employ a computer for all editing operations and make the computer program available to others. (See footnote, page 303, column 2, for a "second-best" solution.)

RATIONALE OF METHODOLOGY

A full discussion of the basic assumptions underlying this general approach to the evaluation of indexing, and of the reasons we handled the data in a given way, is more appropriate for a later report, which will be devoted to the method itself. Two of the major decisions on data treatment, however, require a brief explanation here.

We wanted to measure quantitatively, and as objectively as possible, how well authors, by the indicia they supplied, anticipated the natural viewpoints and terminology of a *group* of scientists. However, at the time we began to analyze the present data, there seemed to be no established techniques suitable for rating, quantitatively, a set of terms one individual had selected by comparing this set with the term-use pattern of a sizable group. In previous studies, an unknown set of indexing terms had been evaluated by comparing it with some standard established by one (or two) individuals; and the findings were expressed as the proportion of terms in the unknown set that "agreed" more-or-less exactly with terms in the criterion set (or sets).

With the present criterion (the term-use pattern of a *group*), this type of comparison was inappropriate. For an index serving a large population, it seemed to follow from our basic premise that a term three-quarters of the scientists in the criterion group used in describing a given document would have a greater probable utility than a term used by only one member of the group, and that some method of weighting should be used to express this superiority. Several weighting procedures were tried before adopting the one reported here, i.e., weighting by squaring the number of criterion group members using a term. The different procedures appeared to give much the same results;* but squaring, a method frequently used to obtain a good "spread" of data, had the advantage of facilitating statistical treatment in later phases of the study and was, therefore, adopted.

The decision to treat each term as if it were independent of any other, i.e., to ignore generic-specific relations among terms, was dictated by the difficulty in preserving objectivity if we had to judge how closely terms are related.

INFLUENCE OF SECTION HEADINGS

If the presence of section headings in the issue of *Federation Proceedings* containing the document sample materially influenced the term-use pattern of the criterion group for a given document, then more members of the group used the SH term than would have had they been responding only to the document itself; therefore, the weighting factor for the SH term is larger than it would have been had the condition of independence not been inadvertently compromised in executing the study design. Since the amount of overweighting that resulted cannot be determined precisely, one must consider whether and how this source of bias ("SH bias") could affect the outcome when the data are used to test the stated hypotheses. The possible effect of SH bias will be discussed in connection with each hypothesis.

AUTHOR INDEXING VS INDEXING DERIVED FROM TITLES

Our hypothesis about the superiority of author indexing over indexing derived from titles was suggested by several years' experience in reviewing and editing the indicia supplied by Federation authors, who often supplied indexing terms that were not included, or even implied, in the titles of their papers.

In using the scores of author and title sets to test the hypothesis, the SH bias favors the author sets, since any overweighting of the SH term represents leading of the criterion group by "exposure" to *one* of the terms in the author set. However, by the same reasoning, the fact that the criterion group was also exposed to document titles, from which *all* the terms in the title set were derived, constitutes a bias favoring the title sets. Since document titles are known to influence strongly the terms individuals use to describe a document, for the purpose of testing this hypothesis, we will assume that the two opposing biases cancel each other. This assumption is conservative since it seems likely that the net effect of the two systematic biases favors the title sets† and is in the opposite direction from the postulated difference between author and title sets.

The data summarized in Table 4 support this hypothesis [compare columns (3) and (7)]. The margin of superiority of author sets over title sets is substantial for each disciplinary group of documents; and for the en-

* We also tried employing the number of criterion group members who used a given term as the weighting factor for that term, and assigning a zero to any term used by only one member of the criterion group. Although the conclusions to be drawn from the present data seem to be little affected by the different weighting procedures we tried, this would not necessarily be true in other applications of this approach to the evaluation of indexing. For this reason, the effects of varying the weighting procedure will be analyzed in the report devoted to the method itself.

† Note that the mean weight of SH terms in title sets is considerably higher than in author sets [column (9), Table 8].

tire document sample, the mean % maximal score of author sets exceeds that of title sets by two-thirds.

GENERAL CHARACTERISTICS OF TITLE SETS

The "efficiency" of terms in title sets, i.e., how much each term contributes to scoring, does not differ materially from that of terms in author sets [compare column (3) in Table 3, and columns (4) and (8) in Table 4]; however, title sets contain substantially fewer terms [column (2) in Table 3, and columns (5) and (9) in Table 4]. The smaller size of title sets accounts for almost two-thirds of the difference between the overall scores for title and author sets.

The SH term was included in less than one-third of the title sets [column (8), part B of Table 3].* Traditional editorial conventions call for document titles to be as brief and specific as possible. Since, in general, the subject categories listed on the printed author indexing form tend to be relatively broad, and since it is from this list that the SH terms were selected, the constraints on the size of title sets and on the "breadth" of terms in titles, together, may well account for all the difference between author and title sets. It is interesting to speculate how much titles might be improved as a source of indexing terms if these constraints were relaxed by a change in editorial conventions.

Relatively few of the terms in title sets are not also in the corresponding author sets; and if the author sets were supplemented by these terms, the mean score of the resulting composite author-title sets for all documents would be increased to only 80%, as compared to 73% for author sets alone [the parenthetical figures, column (7) Table 4, show this increment]. The practical implication is that, given an index compiled from author indicia, the improvement to be afforded by supplementing it with terms derived from titles is rather costly.†

UTILITY OF AUTHOR INDEXING FOR MULTIDISCIPLINARY POPULATIONS

The hypothesis, that authors index better for scientists in their own discipline, originated in a belief commonly expressed by Federation scientists, who emphasized disciplinary differences of viewpoint and language in discussions about appropriate terminology and hierarchical organization for the author-indexing form and about how this form should be revised from year to year.

Tests of this hypothesis would be affected if the section headings in *Federation Proceedings* influenced the term-use patterns of the two criterion group members in the author's discipline either more or less than the patterns of the criterion group as a whole were affected. This possibility appears to be ruled out by the fact that, for author sets, the ratio of the mean score based on disciplinary sub-sets (56%) to the mean score based on full criterion sets (73%) is almost unchanged when the SH terms are *excluded* from scoring (ratio, 37% to 53%).

The present data lend no support to this hypothesis; the observed difference is in the opposite direction from that postulated. This finding has an important practical implication; it means that an index derived from author indicia should serve a multidisciplinary population well—at least when the disciplines lie within the traditional boundaries of the biomedical sciences.

DISCIPLINE-ASSOCIATED DIFFERENCES

The indexing value of titles seems to vary somewhat from discipline to discipline, e.g., compare biochemistry and immunology papers, column (7), Table 4; the difference is significant at the 0.002 level for a 2-tailed t-test. Since this variation is not completely explained by disciplinary differences in the size of title sets [column (2), Part B, Table 3], one might suspect that it is more difficult for authors in certain disciplines to select efficient terms for titling their papers. The data on efficiency of title-set terms [column (8), Table 4] suggest that disciplinary differences of the latter kind exist; if this is true, then constraints on length of titles impose more severe limitations on the indexing value of titles in some disciplines than in others.

A number of studies have assessed the indexing value of titles; but the results of only one of these studies can be compared with the present findings since the others employed a type of criterion basically different from that used here. Bernard and Shilling employed sizable criterion groups of scientists to evaluate how well authors title papers.[12] Their approach and ours were developed independently; but the approaches share enough common features‡ to justify certain comparisons of findings related to titles, especially since the journals they studied represent disciplines that can be broadly classified as "biological" and include two disciplines (physiology and biochemistry) also covered by our study. They reported that their data, an a sample of five articles from each of 10 journals, "suggest that articles in some journals are more difficult to title than those in other journals" and that the articles in the physiology journal studied had better titles than those in the biochemical journal [compare our data in column (7), Table 4].

* Since the SH term represented the author's choice of the one term that best characterized his entire report, the present findings may be interpreted as supporting one criticism of permuted title indexes frequently heard from professional indexers—namely, that titles often do not express the "main" concept of papers.

† In 1963 the decision was made to process document titles for the subject index of the meeting copies of *Federation Proceedings* in addition to the author indicia. In retrospect, we can see that, if the findings for 1962 papers held for later years, processing titles yielded marginal returns. Each author-supplied term contributes to the score of its set an average of 15% [column (4), Table 4]; once the author set had been processed, the incremental return per title-set term would be only 2% [parenthetical figures, column (8), Table 4].

‡ The similarities and differences of the two approaches will be discussed in the report on methodology.

They concluded that disciplinary differences in the adequacy of titles "may be related to differences in the standardization of terminology, to differences in homogeneity of the specialty, and to differences in the diffuseness of articles."

The characteristics of criterion sets for documents in different disciplines (Table 2) are remarkably uniform. Our data on author sets, however, display several disciplinary differences that approach or reach statistical significance, e.g., the differences between pharmacology and pathology papers [in columns (2) and (3), Part A of Table 3, the differences are significant at the 0.001 level for a 2-tailed T-test]; and it is tempting to offer explanations of the observed differences. The present analyses, however, do not provide a basis for determining what part of these differences should be attributed to the various factors that might contribute, such as relative strength of influences tending to standardize terminology, relative speed of development of new research areas, inherent differences in subject matter, etc.

QUALITY OF AUTHOR INDEXING

The origin of our main hypothesis can also be traced to experience in reviewing and editing the author-indexing forms submitted for the Federation meeting; the subjective impression was that authors usually indexed their papers rather well.

Since the SH bias tends to inflate author-set scores, some adjustment of these scores is required in those tests of the main hypothesis where the outcome hinges on how close author-set scores come to the optimal score (maximal score). From the data on the frequency with which criterion group members used SH terms versus terms other than SH terms, one can make a defensible estimate of the *lowest* rating the sample of author indexing could have received under the conditions specified by the study design. We believe it is reasonable to assume that, without any leading by the section heading, the criterion group would have used the SH term *at least as often as any other term* in the criterion set. Under this assumption, the mean weight of SH terms could have been as low as 28 (i.e., the mean weight for the most used term other than the SH term). Using this value, rather than the observed mean weight of SH terms (55, Table 2), one can adjust the scores of author sets given in Table 4 to establish a lower-limit figure of merit for author sets. In testing the main hypothesis, we will use such a lower limit where appropriate.

To test the hypothesis that author indexing is "good," one must ask how high the scores need be to merit this adjective. One way to approach this question is in absolute terms, i.e., by comparison with optimal indexing. On the scale developed here, author indexing averages 73% of the optimal score; the median score is somewhat higher, and almost one-fifth of the author sets score 90% or better, whereas only one-sixth rate less than 50%. When the overall mean score for author sets is adjusted for SH bias, the lower-limit figure of merit is established at 51% of the optimal score. Considering the theoretical difficulties of achieving a score of 100% by this criterion, particularly when generic-specific relations among terms are ignored, can one say that a mean score within the range of 51% to 73% is good? There is another way to make a judgment in absolute terms that is not affected by uncertainty about the proper weighting of SH terms; one can consider what proportion of terms in the criterion sets are included in the corresponding author sets. For all terms in the criterion sets, this proportion averages 33% [column (5), Table 4]; and when terms used by only one member of the criterion group are omitted, it approaches 50%. Do these figures represent good performance? We feel intuitively that, on the basis of either type of comparison with optimal indexing, the author indexing studied here warrants being described as good; however, such subjective judgments are always exceptionable and unsatisfying.

We have not, as yet, collected sufficient data on how other types of indexing score by the same criterion to be able to rate author indexing relative to a variety of basically different indexing techniques, which is our ultimate goal. One is not justified in comparing the authors' performance with that of the individual members of the criterion group, considered as scientist-indexers; the latter would score lower if rated by a similar criterion (mean, 50% maximal score). Chief among the reasons for not considering the data comparable is the fact that members of the criterion group were asked to describe the documents from their own *personal* point of view—if they were functioning as indexers, it was only for themselves. In contrast, the authors knew the terms they supplied were to serve in an index used by a large, multidisciplinary population. They were highly motivated to try, in their choice of terms, to anticipate the different ways others might think about their papers; if they failed, others to whom their work was relevant might not learn of it. In projecting themselves into the position of other research workers, and in addressing the combined interests of a varied group of users, the authors appear to have succeeded remarkably well.

The present data allow one relative comparison of basic indexing techniques, however. Indexes derived mechanically from the words contained in titles are now accepted as useful; and Tukey has suggested that KWIC indexing be adopted as a standard against which other kinds of indexing be compared.[13] Since our findings indicate that author indexing is considerably better than the indexing that can be derived from titles, one could say that author indexing is *relatively* good since it is better than a widely used, hence "acceptable," mode of indexing—a mode that has been

proposed as a standard.*

In a later phase of this study we will assess the indexing of 12 highly qualified, professional indexers by the same criterion used for author indexing and thus obtain comparable data on another major indexing technique. Parallel data on experimental indexing techniques, such as bibliographic coupling and statistical association, would be of great interest.

GENERAL CHARACTERISTICS OF AUTHOR INDEXING

Authors freely exercised the option of writing in additional terms, even though the general relations of their paper to current research could usually be expressed by their choices of the 373 relatively broad subject categories printed on the author-indexing form. By their choices of two subject categories and their write-ins, authors provided a relatively high number of terms per document—almost five terms per document (including zero terms). The complexity of individual terms varies in different kinds of indexes; for example, one of the terms assigned to document No. 100, "Cell, Tissue Culture—Neoplasms" (Table 1), would appear as four separate terms in a permuted title index, and as two or three terms in *Index Medicus.* For this reason, it is difficult to compare the number of subject entries per document ("exhaustiveness") provided by author indexing with the statistics for major bibliographic services. However, author indexing would seem to be at least as exhaustive as that of the broad-coverage bibliographic services.

ZERO TERMS

The terms that were not used by *any* member of the criterion group need further study; almost one-fifth of the terms in both author and title sets were zero terms, even though nonsubstantive terms had been edited out in establishing the sets. Among the possible explanations are the following: the zero terms could (1) be closely related to criterion set terms but be more specific or generic than the latter; (2) describe some aspect of the author's work not explicitly covered in the short document; (3) represent a special language meaningful to the author but of little value for communicating with others. We plan to attempt an analysis of generic-specific levels of the terms in criterion, author, and title sets where such levels can be established objectively. This analysis may indicate what proportion of the zero terms are accounted for by the first explanation.

FURTHER ANALYSIS

The data collected in this study provide excellent material for many types of analyses. We have explored only a few of the possibilities and hope there will be opportunities to collaborate with others in exploiting this resource more fully. (See footnote, page 303, column 2).

SIGNIFICANCE OF FINDINGS

Possible applications of the present approach, and of the general method for assessing the quality of indexing that developed from this approach, will be discussed in a later paper. The major practical implication of the findings related to author indexing may be summarized as follows: if author indexing serves as well as this study would indicate, the demonstrated feasibility and practicability of processing such indexing by computer as a service operation † means that, with author cooperation, subject indexes of good quality, with standardized entries and adequate cross-references, can be produced both rapidly and economically. But whether author indicia are used as input to professional indexers, who edit and refine the product, or as input to computers, the potential for improving the quality, currency, and economy of subject indexes is great.

• Conclusions

1. The data support the hypothesis that author-supplied indicia represent better indexing than terms derived from document titles.

2. For six major disciplines within the broad field of biomedical research, the data do not support the hypothesis that indexing terms supplied by an author serve scientists in his own discipline better than scientists in other disciplines.

3. As represented by the sample used in this study, the quality of author indexing can reasonably be described as "good."

4. The general method for measuring indexing quality developed in this study may prove to be a practical, objective yardstick of wide applicability.

5. The present findings suggest that indexing terms supplied routinely by authors when they submit manuscripts for publication can be used to produce, rapidly and economically, indexes that accommodate well the viewpoint and terminology of scientists in a relatively broad spectrum of disciplines.

• Acknowledgements

This study was made possible by the scientists who served as members of the criterion group and gave generously of their time—Julius Axelrod, Ph.D.; William J. Bowen, Ph.D.; Harry Demopoulos, M.D.; Howard C. Goodman, M.D.; Herman F. Kraybill, Ph.D.; Maurice Landy, Ph.D.; Louise H. Marshall, Ph.D.; Olaf Mickelsen, Ph.D.; Harold M. Peck, M.D.; Frederick

* It should be noted that the title sets we derived from document titles represent considerably more sophisticated indexing than that obtained by simple permutation of words in document titles: therefore, the margin of superiority demonstrated for author indicia would have been materially larger had the author sets been compared with terms in a KWIC-type index.

† For the 1963 Federation meeting, the indexing terms supplied by 3,138 authors were processed by computer programs that standardized terms according to a thesaurus, added cross-references, compiled an alphabetized subject-heading index of the conventional type, and reproduced this index ready for printing directly from the computer printout;[4] 98% of the 29,860 words constituting the input to the computer were handled without human intervention. The total time for human editing was six hours, or about seven seconds per document, as compared with the five to eight minutes per document required in previous years.

L. Rodkey, Ph.D.; Julius Sendroy, Jr., Ph.D.; Earl R. Stadtman, Ph.D.; and Sidney Udenfriend, Ph.D. For criticism and suggestions, the authors are indebted to many colleagues; and, in particular, to a small group which met informally in Philadelphia during the 1964 ADI convention and reviewed an early draft of this paper; and to the Research Fellows (USPHS Grant HE-5414) of the Institute for Advancement of Medical Communication, who supplied continuous "feedback" as the study progressed. Two of the Research Fellows made special contributions—Jeannette F. Rayner performed statistical tests and materially aided in data analysis; and Andrew M. Sherrington, B.M. (Oxon.), edited several of the numerous drafts through which this paper evolved.

References

1. MILTON O. LEE, "Responsibilities for Scientific Information in Biology: Proposal for Financing a Comprehensive System," *Proceedings of the International Conference on Scientific Information,* National Academy of Sciences–National Research Council, Washington, D.C., 1959, Vol. 2, pp. 1417-1428.

2. CLAIRE K. SCHULTZ and CLAYTON A. SHEPHERD, "Scheduling a Meeting and Preparing an Index by Computer," *Federation Proceedings,* Vol. 19, 1960, pp. 682-699.

3. PAULINE ATHERTON, *Aid-to-Indexing Forms,* American Institute of Physics, AIP/DRP 64-2, 1964.

4. CLAIRE K. SCHULTZ, "Editing Author-Produced Indexing Terms and Phrases Via a Magnetic-Tape Thesaurus and a Computer Program," *Automation and Scientific Communication* (26th Annual Meeting), American Documentation Institute, Washington, D.C., 1963, Part I, p. 9.

5. JOHN R. CLARK. "The Author as Indexer," *Science,* Vol. 133, 1961, p. 1040.

6. HENRY BLACK. "Authors as Indexers," *Science,* Vol. 133, 1961, pp. 1634-1635.

7. C. W. CLEVERDON. *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems,* College of Aeronautics, Cranfield, England, 1962.

8. WILLIAM GOFFMAN and VAUN A. NEWILL. *Methodology for Test and Evaluation of Information Retrieval Systems,* Comparative Systems Laboratory, Western Reserve University, Cleveland, Technical Report No. 2, 1964.

9. HAROLD BORKO. "Research in Automatic Generation of Classification Systems," *AFIPS Conference Proceedings,* 1964 Spring Joint Computer Conference, Vol. 25, 1964, pp. 529-535.

10. MARY E. STEVENS and G. H. URBAN, "Training a Computer to Assign Descriptors to Documents: Experiments in Automatic Indexing," *AFIPS Conference Proceedings,* 1964 Spring Joint Computer Conference, Vol. 25, 1964, pp. 563-575.

11. RICHARD H. ORR. "The Metabolism of New Scientific Information: A Preliminary Report." *American Documentation,* Vol. 12, 1961, pp. 15-19.

12. JESSIE BERNARD and CHARLES W. SCHILLING, "Accuracy of Titles in Describing Content of Biological Sciences Articles," *BSCP Communique,* No. 10-63, 1963.

13. DON R. SWANSON. "The Evidence Underlying the Cranfield Results," *Library Quarterly,* Vol. 35, 1965, pp. 1-20.

Toward a National Technical Information System

A. G. HOSHOVSKY and H. H. ALBUM

Headquarters, Office of Aerospace Research, U.S. Air Force, Washington, D.C.

Captain, USAF, Department of Mechanical Engineering, Stanford University, AFIT Program

It is unfortunate, but nevertheless a fact, that our country does not have a working plan geared to setting up a national technical information system. This is not to say that in recent years significant advances have not been made. Perhaps the most notable paper on the problem of handling scientific information is the report of the Weinberg Panel.[1] This report, which was endorsed by the late President Kennedy, stands as a milestone in our progress towards fulfilling the nation's needs for technical information. The Weinberg report analyzed a number of fundamental problems in handling information. It also touched upon the need for improvement in the national network of information handling. The guidelines recommended by this panel are still being used in setting up information programs.

The Weinberg report, coupled with the activities of the Senate Subcommittee on Government Organizations (when it was headed by Hubert Humphrey before his election to the Vice-Presidency of the United States) and the energetic actions of the Committee on Scientific and Technical Information of the Federal Council on Science and Technology, has had considerable impact upon the activities of the Federal government. Practically all agencies of the government have increased the pace of their information activities. At all levels of R&D management we now have a general awareness of the relative importance of technical information, which did not exist before.

The Weinberg report, unfortunately, did not offer a plan for a clearly defined national system for collecting, identifying, distributing, storing and searching for technical literature. This is still the root of our most numerous and perplexing problems today.

Students in this area have long noted that there has been a shortage of general plans which bring into proper focus and give direction to the efforts of individual agencies. These plans are prerequisite to action. Our present "system" has not emerged as a result of overall planning. As the needs arose, individual groups (and agencies) established their own information systems and continued to operate more or less independently. In this respect, we must agree with Hattery [2] and others in saying that we have reached the stage at which such evolution can no longer be a substitute for systematic planning.

Our objective here is to offer a general plan for the construction of a comprehensive national technical-information system. We hope that it will provide the foundation and framework upon which other, more detailed plans can be built. The system we will consider will deal exclusively with the published scientific literature. It excludes the broader consideration of technical documentation such as engineering drawings, technical-logistics data, or classified defense literature. Further, it excludes the even broader aspects of information which rest on personal contacts of scientists and engineers or exchanges of personal correspondence.

This paper consists of two parts. First, we examine other proposals to discover the fundamental factors which prevented their acceptance as a base for a coordinated national technical-information system. We also go back to some of our leading thinkers on the theory of organization to discover the conditions which are required for the emergence of any system.

In the second part we discuss what the national system should be able to accomplish, enumerate the information groups needed, and suggest a general framework within which they can operate. Specifically we

propose that the system be constructed along the functional pattern of information handling (collection, identification, processing, distribution, etc.) to permit each information group to carry out the tasks in which it is most proficient. Finally, we propose a mechanism for controlling the system to insure that it accomplishes the desired objectives.

Some Past Proposals

We have examined a number of proposals advanced in recent years.[3-8] For the purpose of our examination, these proposals (as well as the actual practices of some of our Federal agencies) can be grouped into three general categories. The first category consists of proposals for a "central Federal information agency"; the second encompasses the concept of "delegated repositories"; the third involves strengthening the existing methods and organizations by increased subsidies.

THE CENTRAL STORE CONCEPT

Under the central store concept, a central government agency would handle most (if not all) the published scientific and technical documents. The main part of this system would be a central government depository. As pointed out (but not necessarily advocated) by Alvin M. Weinberg: "A possible *technical* solution to the problem of adequate dissemination of documents is the centralized depository—that is, a central store to which manuscripts are submitted and which distributes copies of manuscripts on request. The central depository would announce its accessions both by title lists and by abstract bulletins. The central depository idea has been proposed innumerable times, and its advantages are impressive: It is extremely fast; it regularizes distribution of reprints; it computes circulating literature; it funnels the accumulation from a given field in one place for efficient retrieval." [9]

There are many advocates of this approach. A bill * for one such center was recently submitted in the halls of Congress by Representative Pucinski.[10] The advocates often contend that this system would eliminate the need for the multitude of similar government agency information systems and give a single point of access for all users, thereby eliminating many of the present difficulties of searching for scientific literature.

So far this approach has not been adopted. There are experts who seriously doubt that such a system would be either really fast or easily accessible. It is even more doubtful that it would eliminate many government agency systems which have peculiar "mission-oriented" needs for documents, many of them highly classified.[11] Its most serious shortcoming, however, is that it would not use the existing nongovernment information resources.

The nongovernment information groups, the professional scientific and technical information societies, are effective communicators and the major source of technical information today. They are an inseparable part of our scientific and technical community. They use their many members to collect, abstract, translate, review, and announce information to an extent that could never be achieved economically by a central repository.

Furthermore, we cannot build a "central store" and hope it will not adversely affect the already weak financial position of private abstracting and information services. With the increased demand for technical information and the rapid growth of interdisciplinary sciences, it is becoming increasingly difficult for the societies to provide adequate financial support to their information services. A new, large government information handling house would, among other things, tend to deprive these societies of their customers. Since the relation between government and nongovernment systems must be of continuing concern [12, 13] the "central store" solution is thus clearly unacceptable.

In considering the national information system we must devise the means of preserving and improving the information activities of our technical societies. We cannot afford to start a new government information system which completely duplicates nongovernment information sources—particularly since it will serve the same people. Any new system which we adopt must be a planned evolutionary step. That is, we must build the new upon the old.

CONCEPT OF DELEGATED REPOSITORIES

Another suggestion was that we divide the responsibilities for handling information in each of the fields of science and technology among the various government agencies. This would move the work closer to many R&D activities than under a central system. Each agency might establish a central repository in its delegated fields.

This scheme would be both cumbersome and inefficient for the following reasons: The division of responsibility would be rather artificial; for who is to say whether fields like chemistry, mathematics, or physics rightfully belong to the Department of Defense, the National Science Foundation, the Atomic Energy Commission, or the National Aeronautics and Space Administration. These agencies would tend to slant their effort towards their mission and might fail to support topics which interest other agencies. Furthermore, the government groups would have to exist in parallel with many similar nongovernment groups. Each agency would have to operate a complete national and worldwide collection program in order to gather complete information in its delegated field. Each agency would have to review a mountain of information to determine which part properly belonged to that agency. This would require large staffs of especially competent people. Each agency would have to identify, abstract and trans-

* Now modified and resubmitted as H.R. 664. It shows considerable departure from the original "central store" concept.

late material in each of its delegated fields. Finally, there is absolutely no assurance that better service would be rendered just because the job was divided among many government groups.

SUBSIDY APPROACH

The underlying philosophy of this approach is for the government to rely completely upon the present information groups of the professional societies and pour more money into their operations. We would then hope that these groups would get together and work out some sort of system among themselves. Of course, the government would help by setting up "coordination" meetings and so forth.

We have followed this concept for the last decade. Although there has been an upgrading of certain information activities, this type of approach has not resulted in what can be called a coordinated national information system. One reason for this is that government agencies which provide the subsidies are not following any "master plan." Each agency and group is essentially acting on its own. Furthermore, we have not worked out the division of roles and missions between the government and nongovernment groups.

By now it should be clear that this approach will not result in a national information system. It is even doubtful if it can prevent the situation from further deterioration. If we continue to follow this concept, the subsidy money will be spent, but we will probably still have the same old problems associated with a disorganized array of information activities.

• Conditions for Emergence of a System

It is unlikely that a national system can be built if we ignore certain preconditions which must exist. Barnard [14] and Simon,[15] as well as Papandreau,[16] point out such conditions which are applicable to the information handling system. They tell us that:

1. The existing groups must recognize a common purpose.
2. They must be willing to contribute actively to a common system.
3. There must be deliberate communication between groups to make the system dynamic.

Most prior plans have stumbled because their advocates have ignored one or more of these points.

For one thing we must agree on a statement of purpose(s). One cannot expect existing groups to develop willingness to cooperate in a scheme where the purpose is general, intangible and perhaps only of sentimental character. Our goals in setting up a national system must be expressed in terms of the roles and concrete outputs of all groups in that system. Unfortunately, here is where we run directly into the often repeated argument that the large number of unknowns precludes setting up a practical national plan.[17] We believe that such an argument evades the issue. The real world is of infinite complexity. We never know all the aspects of most problems before we begin to solve them. As McDonough pointed out,[18] we cannot study forever the problem and seek the one final all-inclusive plan that will be good for all times. We should be able to start now with a rational plan, put it into action, and then modify it as we learn.

Many proposals for a "central store" of information have failed to win support because they, in their design, aimed a direct blow at the pocketbook of existing information groups. It would be sheer nonsense to expect that these groups would be willing to contribute activity to a system which virtually threatens them with extinction. Any plan, to be acceptable, must try to show these groups how they will benefit and how they will fit into the system.

The third condition—deliberate communication in planning and operations—is also something we must have; we don't have it now. We should underscore the word *deliberate,* because there seems to be much communication of the nondeliberate type. There is no commonly acknoweldged meeting ground for all of our information groups, government and nongovernment alike. When many groups are to be linked into a common system such a meeting group is essential to insure good communication (upward as well as downward) so that the overall system is able to function. Any proposal, therefore, must also include provisions for a forum where deliberate discussions can be held and advice from interested parties can be obtained. We can neither define the common purpose nor generate a willingness to establish cooperative effort unless this element is present.

• A Suggested Approach

As pointed out, what we have to say in the following pages is conceptual in nature. We will not discuss such issues as storage and retrieval techniques or methods of dissemination, simply because they belong to the detailed planning. Adequate tools for such planning presently exist.[19] Instead, we will try to deal with the following topics in relation to setting up a national system:

1. What should the system be able to do?
2. What are some of the principles upon which we should base our system?
3. What do we have to work with now, and what additional elements do we need?
4. What should be the relationship between the groups belonging to the system?
5. What controls do we need to assure that the system will operate properly, and how can we set them up?

THE REQUIREMENTS

First, let us look at what our national information system should be able to do. It must put our nation's

scientific research groups into close contact with developments in this country and in the rest of the world. To do this, it must allow for easy and economical access to recorded scientific knowledge regardless of its origin.

Specifically our system should:

1. Collect the whole of the scientific and technical literature from all over the world. This collection process must insure that the literature collected by any one information handling group in the system will ultimately belong to the whole system and will become a part of the total national store of information. This collection process must also include information on research and development which is underway.
2. Sift these items and break them down into subject fields of scientific and technical interest.
3. Index them for future retrieval.
4. Notify our engineers and scientists about the portion of available documents which is relevant to their field of interest and supply them with the required documents.
5. Store all foreign and domestic documents for many kinds of retrieval. It is true that no one individual needs the total store of information; it is equally true that any individual may have a need for access to the total store of information in his defined area. Since such peculiar combinations cannot be easily predicted, an opportunity to search through the total store is vital. This would imply a need for at least one central point, or a switching center, which would make cross-disciplinary searches possible, or for a central point of inquiry about where to get information.
6. Distribute documents (or microcopies) to local service centers (libraries, special information centers, indexing and abstracting services, etc.) according to their stated interests.
7. Provide newly formed abstracting and indexing groups, libraries, centers and special-interest groups with the required collections of previously published literature at a reasonable cost.

THE PRINCIPLES OF CONSTRUCTION

From our previous discussion some of the principles which we espouse should be apparent. They can be summarized as follows:

1. The national system must be systematically planned. There must be communication and agreement on goals, on time schedules and on the means of achieving these goals.
2. The plans for a national system must provide incentives for the present groups to contribute their activity to the network.
3. The system must be established as a partnership of the government and the private sector in pursuit of common goals, and provide for a profitable division of work.
4. There should be no duplication of effort among groups within the national system unless it is deemed necessary and clearly planned that way. In particular, the government should compete as little as possible with the private sector of the national system.

THE ELEMENTS

What resources do we have to draw upon and what do we need? We have the following groups:

1. *Government R&D agencies and their documentation services.* They include such activities as specialized information analysis centers; NASA's Information Facility, Defense Documentation Center, etc.
2. *Government-wide R&D documentation systems.* Included here we have the Office of Technical Services (recently renamed the Clearinghouse for Federal Scientific and Technical Information and placed under the National Bureau of Standards), National Referral Center, National Library of Medicine, Science Information Exchange, and similar entities whose main function is processing of documents or pointing a way to the documents (as contrasted with those entities which handle information as a part of actual R&D activity) in form of consultation and direct answers.
3. *Specialized information centers.* These groups vary widely both in size and the kind of service they furnish. Characteristically they are associated with sizable research institutes where experts evaluate the significance and pertinence of reported work to the subject area covered by the center.
4. *Abstracting and indexing services.* We refer here to the documentation services of both professional and technical societies, and commercial contractors (such as the one which prepares and publishes NASA's abstract bulletin). They are now our major announcement mechanism.

In proposing a system we assume that these four elements would be able and willing to divide the total work along the traditional functions of the documentation process (i.e., the acquisition, production of documents, abstracting, announcing, storage, retrieval, dissemination, etc.). Certain segments of this community have an established capability in one or another of these functions. Thus, we think of the government with its vast network of agencies, reaching practically into every corner of the globe, as best suited to insure an orderly collection of scientific papers and reports. We look upon professional societies and their documentation services as best suited for the task of abstracting, indexing and announcing the profession-oriented literature. Finally, we view the technical libraries and specialized information centers as the "retailers" providing personalized service to their customers, and the

main (but not the only) access to the whole of technical literature.

The above are discrete groups and are best suited to carry out the decentralized tasks. But some aspects of handling the information for an entire nation are best done centrally. For this purpose we need a new element (Fig. 1), a central acquisition and switching group,* which we shall refer to as the National Technical Information Agency (NTIA). Because of the cost involved and the need for a global collection network we feel this is where the government best fits into the program. It could be created from the presently existing organizations such as the Clearinghouse, Science Information Exchange (SIE), National Library of Medicine (NLM), and Library of Congress. Ideally, these organizations should operate under a central control, if such be possible. As a minimum, these organizations must accept clearly delineated areas of responsibility as required by the system. As a whole, the NTIA's function would be to act as the national collector, clearinghouse, central reference facility and supplier of documents to the decentralized private services. Figure 1 illustrates the suggested components of NTIA.[21] Additionally, it can be used as a major platform for experimentation with new techniques and approaches for the progressive evolution of the system.

Now let use examine what the various information

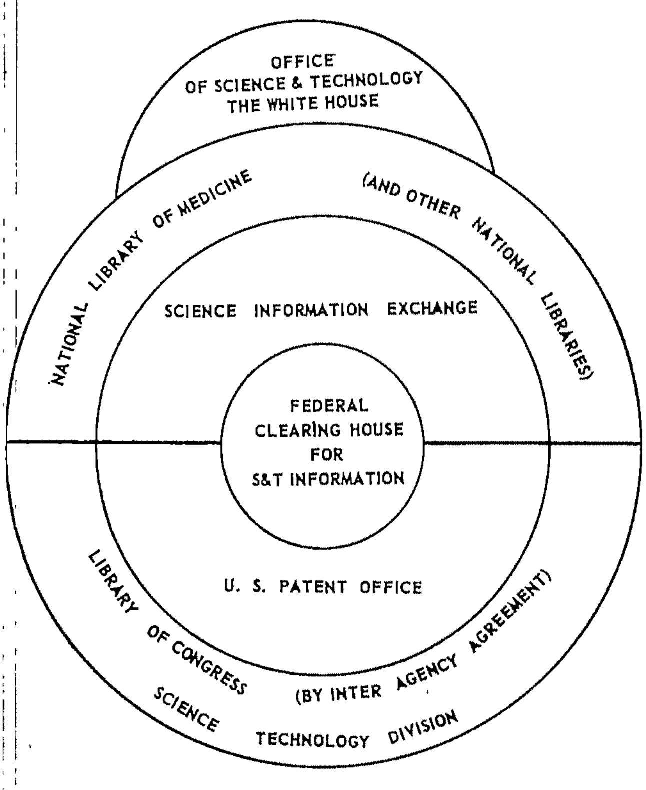


Fig. 1 National Technical Information Agency

groups would do, and how they would relate to each other. Figure 2 illustrates the functional relationship among the groups.

Collection. The National Technical Information Agency would be the focal point for the entire collection effort. It would receive copies (or microcopies for all *new* documents) gathered by the system. NTIA would collect from government agencies, private sources (such as American corporations, universities, and individuals), foreign research and development groups, foreign governments, and from the many information groups which would be associated with the national system.

Information on current technical efforts of an unclassified nature could be fed to NTIA by all government agencies and cooperating private concerns. NTIA would actively encourage the private concerns to list their current research and development projects.

The existence of NTIA would not prevent any group from an independent collection effort. Separate research groups, working on similar projects, would continue to exchange all of their publications directly. Individual engineers would similarly exchange their papers.

It is also clear that in order to set up needed communications, each agency must be free to enter into direct exchange agreements with domestic and foreign research and development groups. Furthermore, foreign groups are more likely to enter into such agreements if the arrangements are made by counterpart working groups. Foreign exchange agreements, however, should be cleared through the National Technical Information Agency (NTIA). This would allow our government to set up a well-planned, consistent, foreign information exchange program. But the major point is that the government agencies would also send NTIA copies (or microreproductions) of all foreign literature which they collect, including translations. Thus our total foreign intake would automatically become available to all other information groups in the United States.

The associated abstracting and indexing services would continue to collect information in this country and abroad. However, now they would be asked to inform the NTIA of their holdings and provide copies or microreproductions of all documents which are not in the NTIA registry.

Identification. NTIA's role in the process of identifying the documents should be restricted to the assignment of the "national accession number," preferably of the type which would contain information on the author, publisher and sequence of issue. The more detailed identification, such as abstracting (when author abstract is not available) and indexing, should be the function of the associated abstracting and indexing groups. Initially NTIA would simply identify the docu-

* A concept of a switching center, developed in detail by Goldberg[20] with respect to an engineering data system, might also be useful in the area of scientific literature.

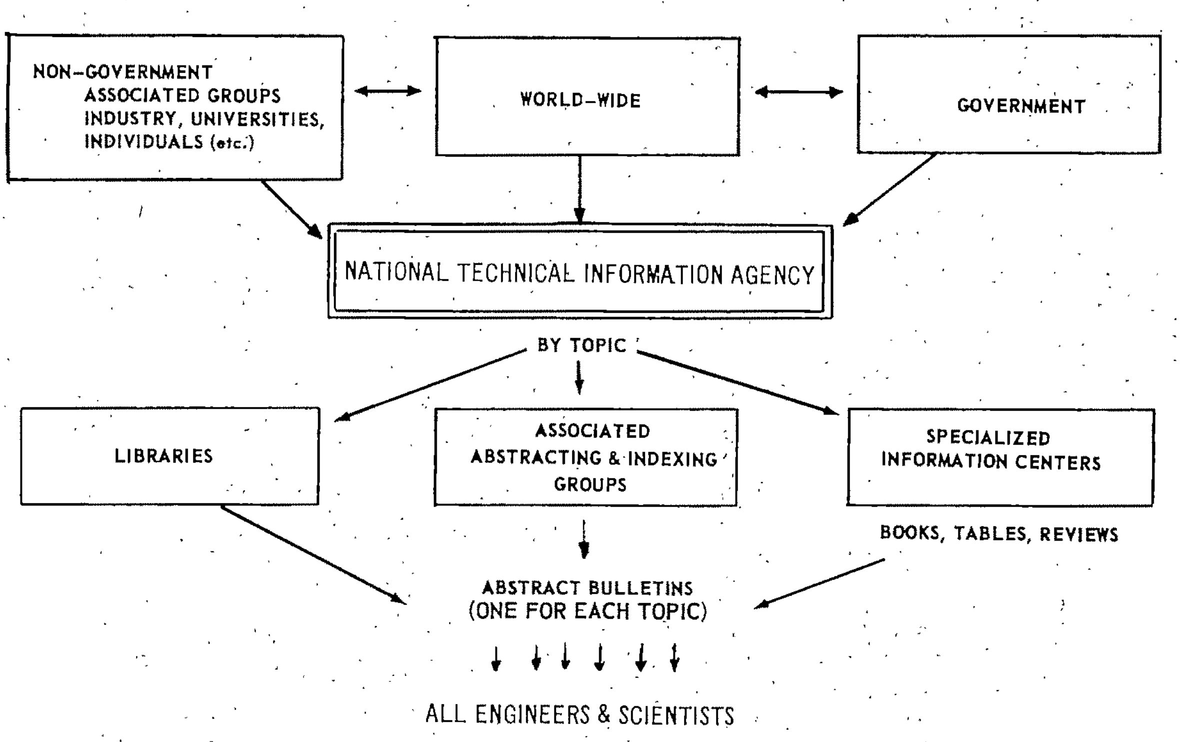


Fig. 2 The Flow of Information

ments in sufficient depth to determine the associated services requiring copies to serve their particular community. NTIA would also contact with selected services for necessary indexing and abstracting work.

Later NTIA would index the material for cross-disciplinary machine searches as may be required by the various information groups, both in government and in industrial concerns. Indexing inputs would be supplied by the associated indexing services.

We believe this type of arrangement would have several advantages. It would overcome the perennial problem of attracting scarce indexing personnel for government work. It would reduce the number of times the same document must be abstracted and indexed. Above all, it would provide the presently hard-pressed information groups with an additional income, so that they may improve their services to the nation as a whole.

Distribution. The members of the associated abstracting and indexing groups would be the primary distributors in the national information system. These groups would publish abstracts, bulletins and bibliographies. Each group would be induced to issue at least one abstract bulletin covering a single field or topic area. Engineers and scientists would seldom have to look further to serve their own areas of interest. The bulletins should be broken down into reasonably sized topics so that each engineer and scientist can afford to purchase the bulletin in his topic area. The same bulletin would serve both government and non-government engineers and scientists. Each of the associated groups would supply its reader with copies of the originals upon request. NITA should pay them for supplying the documents to government contractors.

The various government research and development agencies would handle only the initial distribution of their unclassified technical publications. They would go to the individuals and groups who are known to be concerned with the specific topics covered in the documents, and to the NTIA. NTIA should advise the government agencies on this primary distribution.

NTIA would be the primary distributor to the libraries and specialized information centers and associated abstracting services. These would be supplied by field of interest, topic area, or a combination of suitable criteria. When a library wishes to open or expand its document collection in a new area, NTIA would either supply the needed information or tell the library exactly how to get it. A new specialized information center would receive its information in the same manner. NTIA would also arrange for second printings of government documents for distribution through the associated groups.

Storage. A copy of everything which enters the national system should be on hand at NTIA. Each of the associated abstracting and indexing groups should have complete storage of all available information in its field or topic area. The individual scientists and engineers could then request the references in those areas from the appropriate member. Government agencies would store normal accumulations of scientific and technical information in their libraries and in their specialized information centers. Libraries should be the main source of commercial technical books and primary reference materials for government engineers and scientists. Government and nongovernment agencies alike would normally draw most of the information they need from the associated abstracting and indexing groups. However, when quick-response literature searches are required, or when a cross-disciplinary search is needed, NITA could easily provide such a service.

Searches. Since the individual government agencies would no longer need to operate complete information systems, they would be able to devote themselves more effectively to fulfilling the special needs of their own people. They could each concentrate on mechanizing and operating local quick-response literature search centers. These centers would be under the complete control of the individual agencies to suit the special needs of each agency.

Government search centers would draw their information from the total national collection of scientific and technical literature. NTIA would code the total collection for automatic data processing. The government agencies would have mechanized search centers to use the coded information.

The various agencies would mechanize and staff their individual centers to whatever extent required to meet the demands of their particular mission. Some agencies, such as the Department of Defense, will have to search many fields with a response time of just one or two days. Others, like the Department of Health, Education and Welfare, will have to concentrate on only a few fields, such as medicine and biology. Of course the associated groups having a complete store of information in certain areas are also likely places for individuals or groups to conduct literature searches. Private concerns and government agencies may find it more convenient and less expensive to finance searches at the associated abstracting and indexing group level.

NTIA would have a literature searching service for government groups which do not maintain their own. This service could also be made available to private groups for a fee. Remember, they would now be able to search the entire national store of scientific and technical information (Fig. 3).

In the area of information on current research efforts, NTIA would develop and operate a highly sophisticated system to search out and identify efforts on various scientific and technological topics. This service would be available to managers in government and private enterprise.

Libraries, government and nongovernment, would continue to carry on their traditional role. They would continue to buy and subscribe to the scientific and technical journals. They would be the only centers where scientists and engineers could conduct searches in the area of book literature. The national information system would furnish abstracts of books but would not distribute copies of them.

CONTROLS

Our suggestions are predicated on the assumption that the system will function most efficiently if the individual groups have a high degree of control over their own activities. In fact, the nongovernment groups will operate very much as they do today. They will, however, be asked to divide the work, arrive at common standards, and contribute some of their effort to the common system. Under these conditions we need controls to insure that the system will be able to function and adapt itself to new demands.

As things stand today, the control of government information systems comes largely from the President's Office of Science and Technology and its Committee on Scientific and Technical Information. In the private sector there are groups such as the Federation of Abstracting and Indexing Societies, the Special Libraries Association, Engineers Joint Council, etc. Somewhere in between we have the National Academy of Sciences and its Office of Documentation which is our national representative in the International Federation on Documentation. Finally we have a myriad of government organizations which independently exercise considerable influence over private information activities through their own operations and direct subsidies of many private information activities.

The problem is to unify these varied elements into a structure which would represent the varied interests and provide a mechanism for arriving at agreements. We believe a joint government–nongovernment National R&D Information Policy Council could fulfill this role (Fig. 4).

The membership of the council should come from those groups which already have a substantial amount of scientific and technical information under their control. For example, they might come from the Engineer's Joint Council, the American Institute of Aeronautics and Astronautics, the American Chemical Society, American Documentation Institute, the Special Libraries Association, the Department of Commerce, NASA, etc.

This council would fulfill a number of important roles. It would work to establish standards, modify the planned structure of the system to meet the needs of the times, and see that the government's part in the system is compatible with the needs of the private sector.

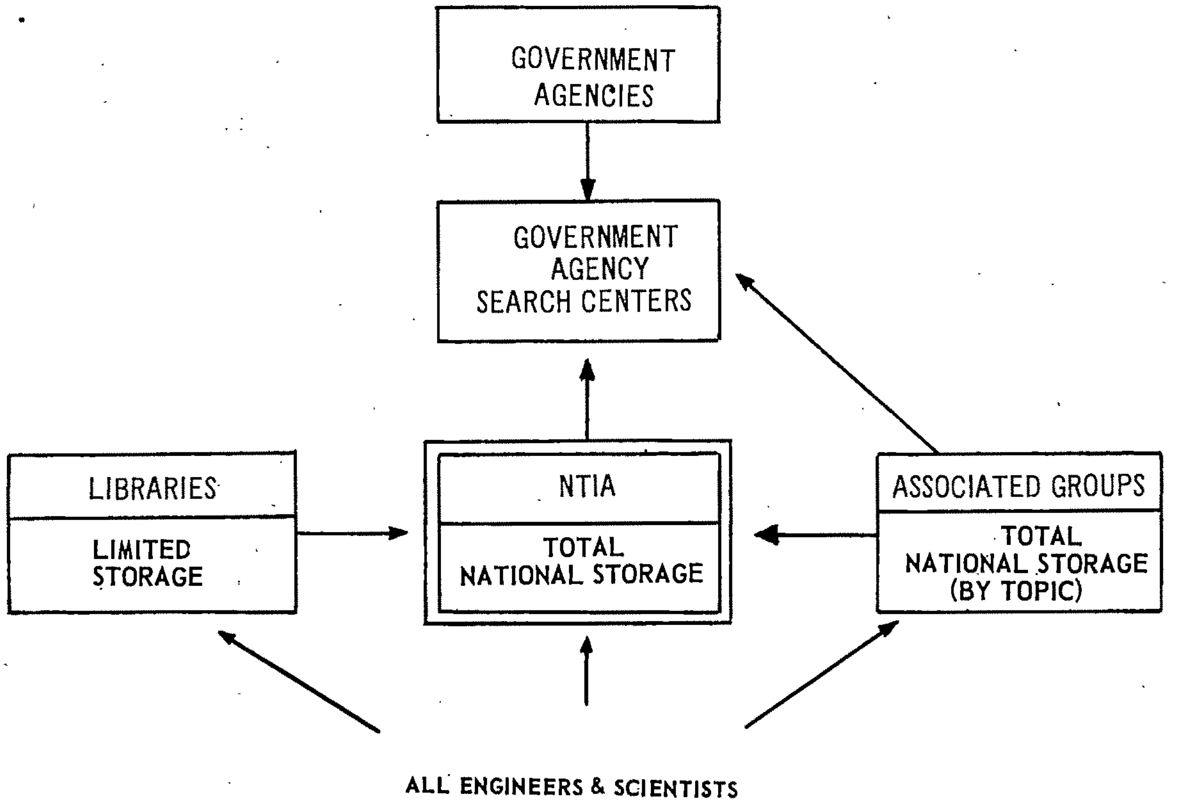


Fig. 3 To Search for Information

• Summary

In the preceding pages we have presented some fundamental factors which inhibit the efforts toward building a truly national technical information network. These factors are:

1. The lack of acceptable plans to provide blueprints for the future.
2. The fear of present groups of losing the autonomy of operations.
3. The absence of a conscious cooperation between the government and the private sectors of the information-handling community.

We outlined a conceptual framework which, we believe, could form a basis for an acceptable information system. The heart of our proposal is the division of work among the government and nongovernment groups. It provides for orderly collection of all scientific and technical literature, a central switching mechanism (the National Technical Information Agency), and the decentralized handling of abstracting, indexing and distribution functions. Storage is carried out both centrally and decentrally to provide maximum access to the information. This would provide access to the total store of technical literature in any desired area and at any level in the system. Further, we proposed a joint government/nongovernment policy mechanism through which to control the performance of the system. We believe the system can be made to work efficiently and, at the same time, give everyone a fair role in it.

There is evidence that the Committee on Scientific and Technical Information has already started serious thinking about the national system. Its newly created Task Force on National System(s) is approaching the problem with remarkable vigor. We hope that what we have had to say here, coupled with the reactions of our readers, will help them in their efforts.

• Acknowledgements

The authors thank the following people: Prof. S. J. Kline (Stanford), Dr. W. E. Teller (Lawrence Radiation Laboratory), Lieutenant General Ely (Committee on Scientific and Technical Information), Mr. Walter Carlson (Department of Defense), and about 20 other people both in and out of government. Through their

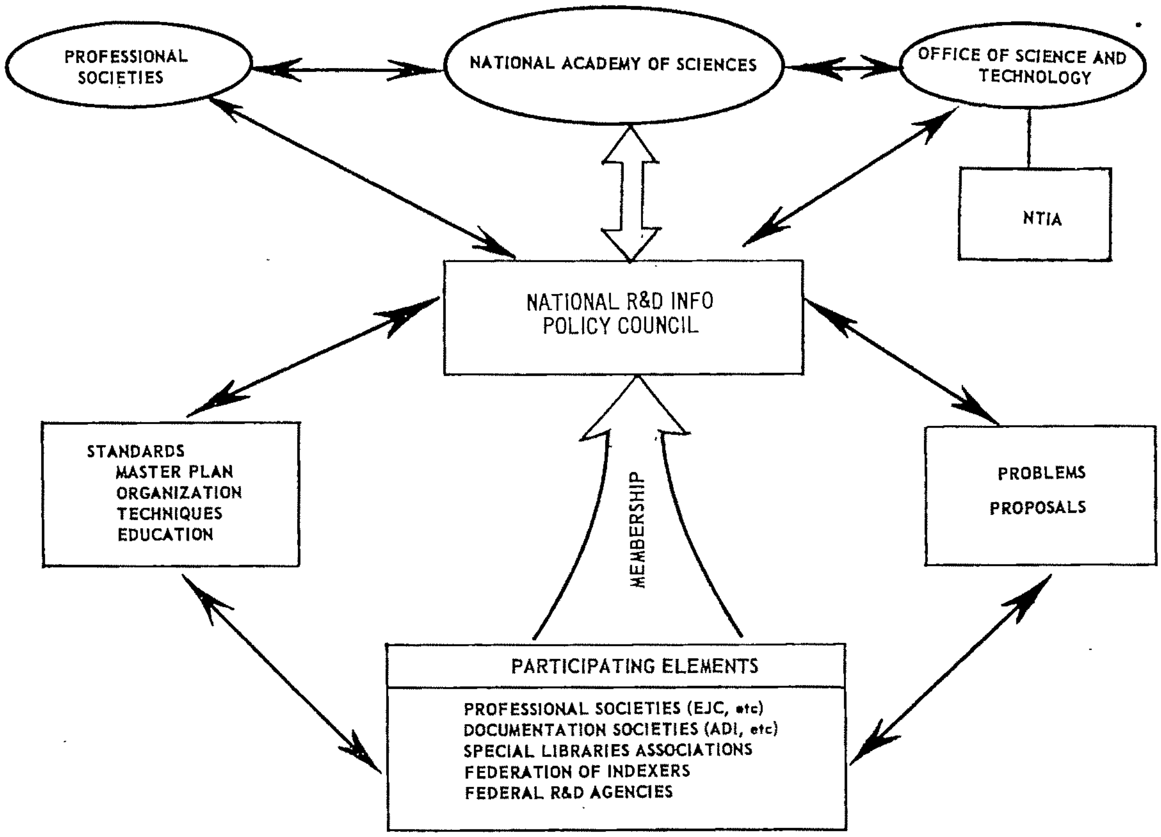


Fig. 4 U. S. National R&D Information Policy Structure

correspondence and conversations they provided helpful suggestions and encouragement. Special thanks are due Colonel A. A. Aines, Executive Secretary of the Committee on Scientific and Technical Information, for his thorough review and constructive criticism.

References

1. The White House, President's Science Advisory Committee, *Science, Government and Information,* January 10, 1963. (Known as Weinberg Report; available from Government Printing Office.)
2. Lowell H. Hattery, "Organization for a Science Information Service," *Information Handling: First Principles,* Paul W. Howerton, ed. (Washington, D.C.: Spartan Books, 1963) p. 173.
3. Julius N. Cahn, "BITS—Bureau of Information for Technology and Science," *Proceedings of the Engineering Information Symposium,* Joint Engineers Council, New York City, January 17, 1962.
4. ———, *A Draft Program for a National Technical Information Center,* Stanford Research Institute, January 1958.
5. ———, *A National Scientific and Technical Information System,* Management Technology Inc., April 1963.
6. G. S. Simpson, Jr., *A Pentagon of U.S. Scientific and Technical Information and Data Service,* a working paper. Battelle Memorial Institute, Columbus, Ohio, November 1963.
7. ———, *Proposal to Establish a Government Corporation to Create and Provide Services from an Integrated Store of Scientific and Technical Information,* Documentation Inc., Bethesda, Md., July 1963.
8. Stafford Warren, *Proposal: The National Library of Science System and Network for the Published Scientific Literature,* Office of the Special Assistant to the President for Mental Retardation, Washington, D.C., May 1964.
9. Alvin M. Weinberg, "Scientific Communication," *International Science and Technology,* April 1962.
10. U.S. Congress, *A Bill to Provide for Science Information Data Processing Center,* H.R. 1946, 88th Congress, 1st Session, January 17, 1963; resubmitted in 89th Congress, 1st Session, as H.R. 664.
11. Andrew A. Aines, Executive Secretary of the Committee on Scientific and Technical Information—an interview.
12. William T. Knox, "Observations on the National Scientific and Technical Information Problem," *Proceedings of the First USAF Scientific and Technology Infor-*

mation Conference, September 30-October 4, 1963, Dayton, Ohio, Office of Aerospace Research, Report OAR-015 (available from Clearinghouse for Federal S&T Information, AD 450 000).

13. THE WHITE HOUSE, PRESIDENT'S SCIENCE ADVISORY COMMITTEE, *Improving the Availability of Scientific and Technical Information in the United States,* December 7, 1958 (attachment to press release of the same date).
14. CHESTER I. BARNARD, *The Functions of the Executive* (Cambridge: Harvard University Press, 1962) p. 82.
15. HERBERT A. SIMON, *Administrative Behavior* (New York: Macmillan, 1947) p. 160.
16. ANDREAS G. PAPANDREAU, "Some Basic Problems on the Theory of the Firm," *A Survey of Contemporary Economics,* Bernard F. Haley, ed. (Homewood, Ill.: Richard D. Irvin, Inc., 1952) Vol. II, p. 186.
17. THE WHITE HOUSE, COMMITTEE ON SCIENTIFIC AND TECHNICAL INFORMATION, *Recommendations to the Federal Council for Science and Technology, Approved 11 Feb. 1964* (unpublished).
18. ADRIAN M. MCDONOUGH, *Information Economics and Management Systems* (New York: McGraw-Hill, 1963) p. 65.
19. MARK CHODROW, DAVID E. SPARKS and DAVID P. WAITE, *Information Service System Modeling: Analytical Tools for Management Evaluation,* Final Report R-1061-4 to National Science Foundation (Contract NSF-C-263), Information Dynamics Corporation, Wakefield, Mass., December 1963.
20. STANLEY A. GOLDBERG, et al, *Engineering Data and Information System (EDIS): Concept and Action Plan Report,* Report No. EDIS-1, Office of Chief of Research and Development, Department of the Army, July 1964 (available from Defense Documentation Center, AD 444 700L).
21. H. H. ALBUM and A. G. HOSHOVSKY, *Threat of Communist Technology: Proposals for Free World Cooperation in R&D and Information Transfer,* Office of Aerospace Research, Report OAR-018, Washington, September 1, 1963 (available from Clearinghouse for Federal S&T Information, AD 430 598).

Indexing Process Evaluation

CHARLES L. BERNIER

The National Library of Medicine, Bethesda, Maryland

The process of subject indexing is difficult to describe and evaluate. The indexer provides guides to subjects reported to authors. The result is an index to subjects—a subject index. He selects subjects, paraphrases and chooses guides to them, and translates the guides into standard index language. The concept of "subject" is elusive. The author controls the number of complexity of subjects and, thus, the indexing density. Guidance among index entries is by cross references. Qualifications of indexers are education in the subject and training in indexing. Differentiation between generalization for brevity and that warranted by the author is important in indexing to the maximum specificity. Weighting of terms in manipulative systems enables increased selectivity upon retrieval. The most common rules for indexing are listed. Teaching and mechanics of indexing are described. Alternatives are contrasted with subject indexing. Indexer-computer collaboration can lead to improved indexing.

The process of subject indexing is best understood by one who has become proficient in it. Yet, a subject indexer, even one of long experience, may have difficulty in describing the process. The following presentation, based upon more than 20 years of experience by the author as an indexer as well as upon the views of many associates, is an attempt to analyze, describe, and evaluate the process. Contrast with other kinds of indexing is used to define and evaluate subject indexing.

The indexer provides guides to subjects reported by authors. The result is an index to subjects—a subject index.

The subject indexer:

1) selects subjects reported by authors,
2) paraphrases the subjects or uses the authors' expressions,
3) chooses guides to the subjects, and
4) translates the guides, if necessary, into standard index language.

The concept of a "subject," an elusive one, is difficult to define and teach. The author's gist, theme, and what he is writing about are other ways of expressing "subject." Items of rational communication, including conversation, telegrams, debates, scientific papers, conferences, movies, technical reports, and books, have subjects. A person can usually select a subject of an item he understands, can paraphrase it, and record it. He may not be able to tell how (by what process) he selects and paraphrases the subject. People can do things without being able to tell precisely how. A famous novelist may not be able to describe in detail how to write a best seller; a cyclist how to avoid falling; or a linguist how to paraphrase every subject in detail, including diction, before it is presented to him.

Subjects, and their number, are entirely under the control of the author and not of the indexer. This is an important point because limitation by indexer, publisher, or others of the number of subject index entries permitted per document will reduce or eliminate indexing of some subjects.[1, 2]

Indexing of technical periodical literature is usually of novel, emphasized, or extensively reviewed subjects because the old subjects were indexed when they were new and do not require reindexing.[3] The user does not want to be referred to what he already knows.

A subject may be compound and separable into two or more simple subjects. An example is "synthesis of acetone and propane," which can be separated for indexing into, "synthesis of acetone" and "synthesis of propane."

Titles or subtitles of technical papers may sometimes

express subjects adequately for indexing. Often titles may generalize solely for sake of brevity, for more rapid and comprehensive understanding, or to aid memory. Titles of patents may be vague. Titles of some reports are unrelated to the subjects dealt with. Titles inadequate for indexing require paraphrasing of textual information. The indexer states in his own words, or in words borrowed from the author, the subject that is to be indexed. The mental paraphrase is made as complete and brief as possible. Paraphrased subjects from the body of the technical articles are nearly always necessary in addition to title paraphrases. Deficiencies of titles for subject indexing are discussed a little later in this paper.

Terms in the paraphrase (or synonyms of them) will, in most cases, be found useful as guides to the subject. Other terms in the paraphrase will be found not so useful as guides. The number of terms required as guides depends upon the complexity of the subject. Complexity of a subject is controlled by the author. It is important to note again that indexing density (the number of index entries per document) is effectively under control of author and not of indexer nor index publisher. That is, the indexer cannot add or subtract subjects, or alter their complexity without overrepresenting, underrepresenting, or misrepresenting what the author is reporting.

The guides to subjects are index entries in a published index, or descriptors in a mechanized index. In the latter, secondary descriptors, roles, or links may also be chosen to restrict the meaning of, or define, the principal descriptors chosen.

The indexer translates, if necessary, the guides chosen from the paraphrase into standard index terminology. This avoids scattering of similar entries. Scattering is a serious fault of poor indexes.[4] Use of synonyms as index headings is the principal cause of scattering. The synonym in the paraphrase is translated into the standard term used as subject heading in the index. The preferred terms are substituted for synonyms frequently in indexing for science and technology. **"Vitamin B_1"** is replaced by **"Thiamine," "Gibbs' function"** by **"Free energy,"** and **"Nomograms"** by **"Nomographs."** Another frequent substitution in manipulative indexes is of more general terms for specific ones. Thus, **"Toyomycin"** may be indexed under **"Antibiotics," "Biosynthesis"** under **"Metabolism,"** and **"Photodynamic action"** under **"Light."** The use of more general terms to represent more specific terms is normal and necessary in indexing with vocabularies of limited size. Such use does not constitute unwarranted generalization because both indexer and user have only the more general term available in the vocabulary and so realize that the more specific term is included under the more general for sake of vocabularly size limitation and generalization.

The translation into systematic nomenclature of chemical organic compounds by authors is difficult. However, it is a skill that is vital at present in preventing scattering in chemistry indexing.

Scattering of like information may also occur among "modifications" (modifying phrases) under subject headings because of differences in paraphrasing. Hence, standard paraphrase rules are used. The indexer translates his paraphrase, when necessary, into a standard form by rule.

The novel, emphasized, or extensively reviewed subjects are identified for indexers by authors in their technical writings. The technical author knows what is new in his work and is usually careful to point it out. In most cases, the author clearly indicates novelty. Lack of novelty in a subject reported is often inferred from specific reference to earlier documents. New subjects are more readily apparent to indexers trained in the subject field of the document than to those untrained. Indexers untrained in a subject field must take time from indexing to learn, bit by bit, the subject field.

The number of simple subjects in articles resulting from chemical research may vary from one to more than a thousand. Experience with *Chemical Abstracts* gives the average as two to four.* Omission of subject-index entries causes loss of access to the subjects through the index. The effect, so far as use of the index is concerned, is as though the author had failed to report the subjects or had not done the research leading to them.

Simple subjects may be combined by indexers into compound subjects under certain conditions. This is done to save space in indexes. For example, the title of an article might be, "Thermodynamic Properties of Twenty-Seven Hydrocarbons." This generalizing title expresses a compound subject. One simple subject derived from the article might be **"Free energy of pentane."** To save index space, the paraphrase might be **"Free energy of hydrocarbons"** or **"Thermodynamic properties of pentane."** The subject-index entries derived from these paraphrases would be, **"Free Energy, of hyrocarbons," and "Pentane, thermodynamic properties of."** The entries, **"Thermodynamic properties, of pentane"** and **Hydrocarbons, free energy of"** would be avoided as unwarranted generalizations unless, of course, such generalizations were made in the article by the author.

Titles of medical articles, for example, usually generalize.† The indexer differentiates between the warranted generalization by the author and the generaliza-

* From experience at *Chemical Abstracts* and the Defense Documentation Center, the number of index entries per document exceeds the number of subjects because, usually, several entries are needed to cover a subject adequately.

† A sample of titles scanned in *Index Medicus* indicated that 51% employed some generalization.

tion in titles for sake of brevity, understandability, or memory. The indexer examines the article to distinguish between the two kinds of generalization. Titles are usually found to be inadequate for complete subject indexing. About 66% of the subject-index entries for abstracts other than organic ones in *Chemical Abstracts* are derived from the bodies of abstracts rather than from their titles.[5] For organic abstracts, the percentage would be higher. About 51% of the titles published in *Index Medicus* generalize. Generalization for sake of brevity requires specific indexing derived from the bodies of articles or abstracts. Titles of technical articles often omit indexable subjects somewhat peripheral to the principal subject of the paper. For example, a paper on adsorption of hydrocarbons by quartz may describe a new balance for measuring adsorption; or a paper on chromatography may describe, in detail, a new adsorbent developed by the author. In the field of science, a few titles are ambiguous or inadequately informative. Titles of patents are often intentionally vague. Titles of unpublished reports may be completely uninformative, e.g., "Progress Report Number Two." Titles of engineering or economic articles are sometimes ambiguous, e.g., "Steels for Automatics," or "Oil Prices in 1965" (from a cottonseed oil journal). Some titles are "cute," paronomastic; some are riddles; and others are "strip-tease," stripped of information so as to tease the reader to continue. However, by far the most serious problems with titles are the frequent use of unwarranted generalization for sake of brevity, and the omission of reference to all indexable subjects.

As the indexer gains experience, he tends to paraphrase subjects directly into standard index terminology. This saves the translation step (4) above. Paraphrasing is a natural art in good conversation. Subjects are communicated compactly by it. It is used to check on accuracy of understanding. The hearer paraphrases what has been said to evaluate both his and the speaker's understanding. The indexer develops compact paraphrasing to a fine art. He treats verbosity as wasteful and confusing. Redundancy in indexes is held to a minimum.

Once a simple subject has been accurately and briefly paraphrased, choice of words in the paraphrase that provide good guidance to the subject is sometimes obvious. "Corrosion of Copper" is an example of a paraphrase for a subject. It is obvious that both **"Corrosion"** and **"Copper"** are two good guides to the subject. Two index entries would be, **"Corrosion, of copper,"** and **"Copper, corrosion of."** The terms used in a manipulative index (e.g., a computerized information retrieval system) would be **"Corrosion"** and **"Copper,"** or their equivalents. The subject paraphrased as "Corrosion of Copper by Acetic Acid" has three good guides: (1) **Corrosion,** (2) **Copper,** and (3) **Acetic acid.** The published index entries corresponding to these three guides are reasonably apparent.

Besides guidance to subjects, the indexing process also provides another kind of guidance. It is guidance to the "right" term when the index user looks in the "wrong" place first. Cross references are used [4]: **"Vitamin B^1,** See *Thiamine.*" **"Photodynamic action,"** See *Light.*" And **"Rust.** (See also *Corrosion*)." A cross reference is unnecessary when both terms are alphabetically close together. Control over the number of cross references used in indexes is usually exercised by assuming a degree of education of index users. Thus, organic chemists would not need the cross reference, **"2-Propanone,** See *Acetone.*"

The above description is an oversimplification of the indexing process. Many complications occur. Subject identification is difficult to teach and apply, especially when the subjects are complex or implicit. The person without knowledge of the field indexed finds consistent identification of all subjects impossible. Dividing compound subjects into simple ones and combining simple ones into permissible compound ones requires education in the subject field. The possibility that an untrained indexer without education in chemistry might possibly provide effective guides to 50% of the subjects in articles dealing with results of chemical research or engineering should not lead the index publisher to believe that he has discovered a way of making diamonds from glass. It is the last 50% of indexing for such articles that has been found difficult to attain.*

Paraphrasing for subject expression is learned slowly. Even experienced indexers become rusty after a few weeks without practice.

Detection and avoidance of indexing of generalization used for sake of brevity (often done in titles) requires considerable skill, alertness, and subject knowledge. Detection and indexing of generalization warranted by the author requires the same qualities.

Selection of terms from the paraphrase to be used as guides to subjects is nearly always more difficult than in the simple examples given.

When weighting of terms is used as in some manipulative systems, selection of guides is supplemented by selection of auxiliary terms of secondary weighting.[6] These secondary terms are correlated during search with the primary terms to increase relevance of output. A paraphrased subject might be "Carotenogenesis in Relation to Microbial Cells." **"Carotene," "Biosynthesis,"** and the classes of microbial cells (**"Bacteria," "Molds,"** and **"Yeasts"**) are good guides to the subject. **"Microbial cells"** is, let us say, too broad as a guide. The author, we assume, generalized to save space and not to make his

* Beginning indexers at *Chemical Abstracts* often had 50-75% of their index entries changed upon checking. Perhaps half of these changes were of a minor nature (e.g., for paraphrasing, abbreviation, elimination of redundancy) and not of a major nature (e.g., omissions, scattering, mistakes or errors). Had these indexers not received their B.S. or higher degrees in chemistry, then the percentage of changes would have been greater.

work literally include all microbial cells. The paraphrase could read "Carotenogenesis in Relation to Certain Microbial Cells" or ". . . in Relation to Bacteria, Molds, and Yeasts." The word "relation" here is also a generalization for the specific relation(s) studied. A specific relation could have been: "cause," "stimulation," or "inhibition." The paraphrase could better have been worded: "Carotenogenesis Stimulation by Bacteria, Molds, and Yeasts." **"Stimulation"** would not be used as a primary indexing term since the author was not studying stimulation *per se*. Thus "stimulation" is not as useful as a guide to the subject studied as are **"Carotene," "Bacteria," Molds,"** or **"Yeasts."** However, it could be useful as a secondary term in a mechanized system for separating irrelevant studies in which carotenogenesis is, perhaps, "inhibited" or "used." **"Stimulation"** would, of course, be a heavily used subject heading because many things are stimulated in biology. This heavy use would tend to reduce selectivity. **"Stimulation"** could also be correlated, for example, with **"Bacteria"** to separate the subject paraphrased above from irrelevant ones, say those dealing with "pathogenesis," "growth," or "culture."

The indexing process has acquired its collection of rules. Since these have been detailed elsewhere,[7] a few of the more significant in common use for the technical periodical literature are listed here:

1. Choose to index those subjects that are novel, emphasized, or extensively reviewed. (Old material was indexed when it was new. Limitation of indexing to subjects of interest to an organization is also practiced to save money on index entries predicted to be of little use. Such limitation is as valuable as the accuracy of the prediction allows.)
2. Index to the maximum specificity warranted by the author. (Unwarranted generalization by the indexer pushes the author beyond what he has actually studied and reported. This rule does not conflict with use of the more general term to represent the more specific when vocabularies of limited size are employed. This point was discussed earlier. The fact that unwarranted generalization may be practiced in some organizations does not justify broadening of author's subjects.)
3. Choose terms most frequently used in the field as subject headings, subheadings, or descriptors. (More users will look for these first.)
4. Provide guidance among headings and from synonyms. (Cross references are a common form of guidance.)
5. Use "modifications" (modifying phrases), titles, subheadings, or secondary descriptors in addition to the terms chosen as subject headings. (These provide more specific and better guidance, and avoid blocks of undifferentiated reference numbers.
6. Check all index entries. (Subject indexing is an art and not so precise, for example, as author indexing. The cost of lost technical information, due to inadequate or missing index entries, may be far greater than the cost of checking.)
7. If a limited vocabulary is used, arrange for its rule-control, prompt updating and for prompt communication of vocabulary changes among indexers and organizations using the vocabulary.

• Teaching the Process

Instruction in indexing has been done effectively by coaching. The coach follows the indexing and shows, with explanations, all changes to the indexer. Self-learning and classroom instruction have not proven to be very successful.† Apparently the process of indexing is too subtle or complex to be learned well and rapidly in these last-named ways.

• Mechanics

The mechanics of the indexing process varies with the organization. At *Chemical Abstracts*, the indexers dictate entries to save time of writing and to produce cleaner copy for the editor and printer.[8] Because well-trained indexers at *Chemical Abstracts* normally make 5-10% of serious errors ‡ in their work, indexing is checked by another indexer, preferably by one with more experience. Some checking systems involve examination of all index entries in relation to documents indexed. All parts of the entries are checked, with possible exception of the reference, which is sometimes checked by clerical help. Discussion of changes with the indexer enables the two to come to common agreement, which helps to eliminate errors introduced by the checker. Editing of assembled index entries removes errors that have escaped checking. Effective mechanics can reduce cost and speed indexing, as well as improve quality.

• Alternatives

There have been repeated attempts to hinge subject indexing on what information or data users may want. The fact that specific users are unknown (some even unborn) at the time of indexing makers predicted information wants or needs of users tenuous as a basis for subject indexing. Users may know their information wants but it is doubtful that they can fully know their

† The author has attempted to teach indexing in the classroom. The experiment was not very successful owing to the large amount of individual attention needed for a long time. Group instruction has been considered at *Chemical Abstracts* and for *Index Medicus* and the *Technical Abstract Bulletin*, but was abandoned in favor of individual coaching, which has been the method of choice. Programmed instruction for selection of subject index headings was tried at *Chemical Abstracts* but abandoned.

‡ Again, this derives from the experience at *Chemical Abstracts* for indexers with many years of experience in the nonorganic field.

needs, especially for novel developments unknown to them. What novel information or data they need may not be what they want before they find it in a display. This casts no reflection upon competence of users or indexers, since prediction of research results may be necessary in order to forecast actual needs of users. Accurate matching of wants and needs involves prediction or invention.

Other attempts have hinged subject indexing on "importance" of items presented by the author. Such items have included subjects, topics, "ideas," "concepts," titles, subtitles, phrases, and even words. Evaluation of importance may involve not only judgment of author and indexer but also prescience, if future value to science and mankind is to be judged. If "importance" is taken to mean "importance of novel developments to the author at the time he wrote the paper," then the result of indexing on the basis of this kind of importance may be substantially the same as subject indexing. That is, the author will usually emphasize novel subjects and indicate old subjects.

Still other attempts have been made to hinge subject indexing on words, clues, titles, etc. It is reasoned that the indexer sees items to derive subject index entries. So why can't computers examine the same items and use the same rules as the indexers so as to come up with the same subjects that indexers choose? [9] Then, by a standard vocabulary, thesaurus, or term-translation table, have the computer present standard, well-formed index entries untouched by human bias—sometimes magnified to make this approach seem rosier.

This approach may well lead to the most effective man—machine collaboration. The indexer helps the computer and vice versa. The indexer might, for example, identify and paraphrase subjects. As an additional aid to the computer, he could, with but little added effort, indicate, say by underlining, those terms that best acted as guides to the subjects. Then the computer could take over and, with the help of a human lexicographer for new terms, translate terms in the paraphrase into standard index terminology. The index entries could be derived from these standard terms plus the paraphrase as "modification" (modifying phrase). Such modifications often would not be so effective as those developed by human indexers. [10] Also, further computer programs might be developed for organizing and standardizing the paraphrase so as to provide tailor-made modifications for each term used as a guide. Eventually computer programs might be developed to select from the paraphrase the terms that were best guides along with secondary descriptors. Even when the computer picks all non-stopped words from the paraphrase or title, as it does in KWIC (Key Words in Context) indexing, the resulting index has been found used. [11] The increased load of irrelevant entries on the searcher is reported not to be a serious problem. The reason for this is that the process of writing a title or paraphrase eliminates most words that would not be good guides to subjects.

Concordances have been reinvented. Concordance words in titles (CWIT) have been advocated as alternatives or temporary stand-ins for subject indexes. Various ways of improving the product have been developed. The subtleness of qualities of indexes makes survival of concordances possible until improvement sets in. Subtleness comes from the invisibility of nearly all omissions and of scattering of like information. Also a very important point is that a concordance is better than nothing. Partial communication of research results is much less expensive than no communications at all. [12] The defects of CWIT as substitutes for subject indexes are well known to indexers and include:

1) scattering of like information;
2) omission of subjects, and of guides to them;
3) extra entries that are not good guides to subjects;
4) inadequate internal guidance; and
5) unwarranted generalization.

The deficiencies of titles, discussed earlier, are one cause of inadequate CWIT indexes. The uncontrolled vocabularies of authors cause the scattering of like information largely because of use of synonyms and general terms for specific.

Citation indexes function as subject indexes because authors tie subjects together by citations. The point of access here is not an index entry but a paper in the field of interest.

Associative retrieval, which depends upon associations among words, documents of a collection, and questions, [13-17] may depend for its apparent success in demonstrations upon the selectivity of abstracts. This point is being investigated.

• Conclusions

Subject indexes guide users to subjects reported by authors. Neither author nor citation indexes are subject indexes, although they are used to locate and discover subjects in a different way. Indexes to wants or needs of users are not subject indexes, nor are indexes to "important things"; neither are word indexes (concordances) nor indexes based upon clues, associations, statistics, or syntax subject indexes. Whether any of these can substitute effectively for subject indexes remains to be proven. Certainly a number of them are already useful as adjuncts to subject indexes.

References

1. CHARLES L. BERNIER, "Correlative Indexes. IX. Vocabulary Control," *American Documentation*, Vol. 4, 1964, p. 101.
2. CHARLES L. BERNIER, "Correlative Indexes. X. Subject-Index Qualities," *American Documentation*, Vol. 4, 1964, p. 104.

3. Charles L. Bernier and E. J. Crane, "Correlative Indexes. VIII. Subject-Indexing vs. Word-Indexing," *Journal of Chemical Documentation*, Vol. 2, 1962, p. 117.
4. Reference 2, p. 105.
5. Charles L. Bernier, "The Indexing Problem," *Journal of Chemical Documentation*, Vol. 1, 1961, pp. 25-7.
6. Reference 1, p. 102.
7. Reference 3, pp. 120-121.
8. Charles L. Bernier and C. C. Langham, "Dictating Machines as Indexing Aids," *American Documentation*, Vol. 6, 1955, pp. 237-238.
9. John O'Connor. "Correlation of Indexing Headings and Title Words in Three Medical Indexing Systems," *American Documentation*, Vol. 15, 1964, pp. 96-104.
10. Ben Weil. Personal communication about the indexes derived from "artificial titles" and used at Esso Research & Engineering Laboratories, Inc.
11. Reference 2, p. 106.
12. N. E. Maron and J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval," *Journal of the Association for Computing Machinery*, Vol. 7, No. 3, July 1960.
13. H. E. Stiles. "The Associative Factor in Information Retrieval," *Journal of the Association for Computing Machinery*. Vol. 8, No. 2, April 1961.
14. N. E. Maron. "Automatic Indexing: An Experimental Inquiry," *Journal of the Association for Computing Machinery*, Vol. 8, No. 3, July 1961.
15. L. B. Doyle. "Indexing and Abstracting by Association," *American Documentation*, Vol. 13, No. 4, October 1962.
16. Arthur D. Little, Inc., "Centralization and Documentation," Final Report C-64469 to the National Science Foundation, 1964, pp. 48–49.

Attitudes of Scientists Toward a Specialized Information Center*

E. J. FEINLER, C. J. COOK, O. HEINZ, and C. P. BOURNE

Stanford Research Institute
Menlo Park, California

Data were obtained from more than 300 atomic and molecular physicists in the U. S. concerning their use of existing information sources. A discussion of their information needs and preferences is presented. An attempt is made to define the attitude of this group toward a specialized information center in the light of their responses.

Introduction

In the course of a recent investigation[1] concerning a proposed information center for atomic and molecular physics, data were collected by means of questionnaires and personal interviews on the scientist's habits and attitudes toward information handling. This paper presents a summary of the data collected along with some of the background pertinent to our study.

It will help the reader to have some facts about the nature of this particular scientific community. About 1,100 scientists in the United States are estimated to be carrying on research in atomic and molecular physics.† This number constitutes roughly 10% of all physicists engaged in research. Work is carried on primarily by individual scientists or by small groups and is fairly evenly divided among academic institutions, private industry, and government laboratories. There are relatively few large projects involved, as can be gauged from the fact that median cost per research contract is $25,000. An estimated 2,000 documents, of which one-third are published papers, are generated in atomic and molecular physics on a world-wide basis annually. Research qualifies as basic rather than applied; however, there are closely related areas such as astrophysics, plasma physics, and geophysics for which results do have immediate applicability. Historically, this is rather an "old" field which was somewhat eclipsed in the 30's and 40's by the emphasis on nuclear physics. More recently the research volume has risen sharply, stimulated in part by the interest in upper atmosphere and space research and controlled thermonuclear research. Since there exist no specialized journal for this subject, articles are scattered through numerous publications on physics, chemistry, and engineering.

Our study attempted to determine the answers to such questions as: What are the current information needs of physicists in this field? What information sources are available to them? How and to what extent are the available sources utilized? What would be expected of an information center specifically designed for scientists in this field? To what extent are the scientists themselves willing to participate in the operation of such an information center? Some of the replies gave considerable insight into the information habits of this scientific community and are likely to be applicable to other scientific specialties as well.

Of 613 questionnaires mailed out, 343 were returned and 100 persons were contacted through personal interviews. Although there was some overlap between these two groups, the number of responses elicited amounted to more than one-third of the 1,100 scientists in the field.

* This work was supported by the Division of Technical Information, Atomic Energy Commission, Washington, D. C.

† The technical definition of the field was originally suggested by two of the authors (Cook and Heinz), who are actively engaged in research in atomic and molecular physics. It was then modified to represent a composite of the views of all the technical workers participating in this study.

• Use of Existing Information Sources

To determine the abstracting, indexing, or current surveillance sources (i.e., secondary sources) most frequently used in this field, the scientists were asked to rate a list of sources as to use. The results are summarized in Table 1.

Physical Abstracts, Chemical Abstracts, and, to a lesser extent, recently introduced *Current Contents* constitute the bulk of the secondary sources used for keeping abreast of the literature. There were frequent complaints about the slowness of the abstracts and the lack of current cumulative indexes. Abstracting and indexing sources published by government agencies are rarely used and their very existence is unknown to a large portion of the scientists—even though these publications contain information useful to this field.

• Use of Existing Information Centers

The extent to which pertinent government information centers were consulted emerges clearly in Table 2.

Although these centers act as depositories for unpublished reports, periodically disseminate pertinent listings, and are available for literature searches and other types of reference services, only the Defense Documentation Center is regularly used by more than 10% of the scientists interrogated. Not only are the centers not regularly used, many scientists were not even aware of their existence. This would seem to indicate either a lack of effective "advertising" on the part of the information centers, a lack of enthusiasm on the part of the scientists, or a combination of both. Information centers are mostly used by scientists in government or industrial laboratories; university researchers keep almost totally aloof from these sources of information.

• Preferences for Various Types of Information Services

An attempt was made to find out whether there were any clear-cut preferences among the scientists for the different types of services a specialized center might offer. Eleven possible services were listed in our questionnaires and scientists were asked to establish their own priorities. The results are summarized in Table 3.

We believe that the rather inconclusive results shown in this table are due to the fact that most of the respondents had little opportunity to crystallize their thoughts on this problem. It is felt that the replies reflect the current habits and needs of the respondents rather than suggest how an ideal information center should function. This fact became particularly evident during the personal interviews.

• Views on Information Needs

The 100 scientists interviewed were asked for suggestions concerning their information needs and desires. Since scientists work on a broad range of problems and in a variety of institutions, it is hardly surprising that their most pressing needs differ, as well as their suggestions for fulfilling these needs. However, a few ideas kept recurring:

TABLE 1

Use of Secondary Information Sources

	Number of Respondents			
Source	Used Frequently (more than 6 times annually)	Used Rarely (1-6 times or an unspecified number annually)	Never Used	Existence of Source Not Known
Physics Abstracts	112	67	60	4
Chemical Abstracts	53	57	134	2
Current Contents	38	15	159	31
Nuclear Science Abstracts	24	42	166	11
TAB (abstract bulletin put out by DDC)	16	12	139	76
Chemical Titles				
STAR (abstract bulletin	15	21	179	29
put out by NASA)	10	12	140	81
International Aerospace Abstracts	8	12	154	69
Solid State Abstracts	7	17	203	16
Physikalische Berichte	6	12	188	37
Science Citation Index	5	7	158	73
Other	1	3	..	..
Chemisches Zentralblatt	0	12	182	46

TABLE 2

Use of Existing Information Centers

Source	Used Frequently (more than 6 times annually)	Used Rarely (1-6 times or an unspecified number annually)	Never Used	Existence of Source Not Known
	Number of Respondents			
DDC (Defense Documentation Center, formerly ASTIA)	28	42	134	39
AEC Division of Technical Information Extension, Oak Ridge	9	22	169	43
NASA Office of Technical Information	9	20	169	47
OTS (Office of Technical Services) Dept. of Commerce	8	46	137	52
IGY World Data Center	6	9	162	6
National Referral Center for Science and Technology (Library of Congress)	2	9	141	91
SIE (Science Information Exchange) Smithsonian Institution	1	5	125	112
U. S. Patent Office	1	14	201	27
Others	1	3	..	..

TABLE 3

Preference for Information Services

Types of Services	1st, 2nd, or 3rd Choice	4th Through 11th Choice	Omitted
Provide state-of-the-art reports	108	28	107
Provide bibliographies and special literature searches	106	42	95
Abstract current literature	90	22	131
Compile data sheets from literature (on request)	74	37	132
Provide mechanism to insure writing timely review articles	73	45	125
Obtain reprints (foreign and domestic)	68	29	146
Provide translation services	62	35	146
Provide continued surveillance of subject fields	62	33	148
Provide quick response to questions about technical content of current literature	57	32	154
Compile data sheets from literature (provided routinely)	57	22	164
Provide reference files for personal use	53	19	171

1. A major concern was promptness. Most of the researchers interviewed wanted quick response to their queries: they tended to discredit agencies doing custom literature searches due to the time lag involved. Several expressed concern about the promptness of indexes and abstracts of current literature. On the whole they preferred to see a good portion of any information dollar go into keeping current, and expanding, such sources as *Physics Abstracts* rather than starting new information ventures.
2. Many felt that a major problem with current literature searches (especially machine searches) is that they attempt to give *all* the information on a given subject indiscriminately. The result is a large bulk of references or abstracts that are not conveniently organized. The recipient must then sort and evaluate this bulk to find the few items of interest.
3. Concern was repeatedly expressed over the present lack of communication between a literature searcher (be it man or machine) and the scientist requesting the search.

Many scientists felt that information centers and information "specialists" were of little use because they either did not understand the scientific material requested or, if understood, they could not retrieve it to the satisfaction of the scientist. The current machine searching (1964) was considered particularly deficient on this score since most machine systems are not indexed specifically enough. Many scientists felt they should either do literature searching themselves or have a scientifically trained person who "understands the problem" do it for them with close liaison between scientist and searcher. The respondents had particularly strong feelings that no one but a competent scientist could or should evaluate data from the literature.

4. There was also some feeling that a competent, specialized journal in the field of atomic and molecular physics could alleviate some of the present problems caused by the spreading of information among many existing journals.

• Attitudes Towards Establishing a Specialized Information Center

When scientists interviewed were asked how they felt about establishing an information and data evaluation center in atomic and molecular physics, the majority reacted with cautious, and in some cases conditional approval—while only about 10% rejected a specialized center outright. This was confirmed by our questionnaires.

However, even scientists generally favorable to the center tended to be haunted by fears of bureaucratic inefficiency, needless duplication of effort, and a lack of understanding of their particular problems leading to an "indigestible avalanche of mediocre information." Comments were repeatedly made that the centers must be highly responsive to user needs. It was emphasized that ways must be found to organize the center so that it would remain under *scientific* guidance and be closely associated with an active and respected research group; under no circumstances should it simply pour out unorganized lists of references or unevaluated results. Characteristically, some scientists even objected to receiving routine compilations of data or bibliographies in their own field, because their desks were already too cluttered!

While opinions were few and divergent about the best structure and staffing of an information center, it is significant that about one-third of the total number of respondents (questionnaires plus interviews) volunteered to contribute their time for active participation as follows:

Amount of Time	*No. of Respondents*
5% or more	68
Conditionally or less than 5%	56
No time	98

It was suggested that research contracts allocate funds up to 10% for this purpose so that adequate time could be devoted to information processing and technical assistance could be used where needed.

• Conclusions

Our overall impression from this investigation of a small but typical scientific community is that existing communication channels are judged inadequate, even though there is no consensus on alternative solutions. What does emerge from numerous comments, suggestions, and criticisms is, as yet a ill-defined but generally held, concept of a smaller and more specialized operation which would be more closely linked to various phases of research. It should serve a multiplicity of purposes: promotion of certain areas of research; evaluation, compilation, and effective retrieval of data; periodic reviews of results and progress in the field. In adidtion, it must effectively carry on the conventional functions of storing and circulating all pertinent foreign and domestic literature and providing reference and bibliographical services.

As one tries to clarify and define these functions more carefully, it becomes increasingly apparent that scientists themselves must play a major role in staffing and operating any such system. The growing awareness of this new responsibility on the part of scientists manifests itself in their willingness to participate in various aspects of the information process, at least within their technical specialty. It is to be expected that through personal involvement in a specialized information system scientists will gain confidence in the reliability of the system and at the same time will gain familiarity with newer communication channels.

• Addendum

Since the completion of this study, an Atomic and Molecular Processes Information Center has been established at the Atomic Energy Commission's Oak Ridge (Tennessee) National Laboratory to compile and evaluate data on atomic and molecular physics.

Sponsored jointly by the National Bureau of Standards and the Atomic Energy Commission, the center will be part of the National Standard Reference Data program and will serve as a focal point for the collection, storage, evaluation and dissemination of information generated throughout the world. All information will be evaluated by scientists working in the field of atomic and molecular collisions.

Initially the center's activities will be limited to atomic and molecular cross-section data and other particle collision process information in these specific areas: (1)

the interaction of heavy particles, (2) particle penetration through matter, and (3) excitation dissociation, ionization and detachment by external electric and magnetic fields.

References

1. C. COOK, et al, "An Atomic and Molecular Processes search Institute, Letters Project 4828. Final Report, search Institute, Menlo Park, Calif., Project 4828, Final Report, 1964.

BOOK REVIEWS

10/65—1R **Automatic Indexing: A State-of-the-Art Report.** 1965. Mary Elizabeth Stevens. National Bureau of Standards, Washington, D.C. (NBS Monograph 91.) Superintendent of Documents, U. S. Government Printing Office. 223 pp. $1.50.

This report is one of a series sponsored by the Research Information Center and Advisory Service on Information Processing. RICASIP is jointly supported by the National Science Foundation and the National Bureau of Standards, and is engaged in a continuous program of collecting information about research and development activities in the field of information processing. The report is intended to be a state-of-the-art survey of "current progress in linguistic data processing as related to the possibilities of automatic mechanized indexing" to improve "cooperation in the fields of information selection systems development, information retrieval research, and mechanized translation." (Introduction) Coverage of the literature is through February 1964.

For the purpose of the report, automatic indexing is defined as the use of machines to extract or assign index terms without human intervention. The actual scope of the report is somewhat broader than that which this definition would imply. "Machine-like" indexing by people was considered as falling within the scope of automatic indexing, together with such related areas as evaluation methods and the machine preparation of card catalogs, book catalogs, and bibliographies. The literature on the machine generation of indexes is grouped under the following headings: automatic derivative indexing, automatic assignment indexing, and automatic classification. The report also covers thesaurus construction, statistical association techniques, probabilistic indexing, natural language text searching, research in linguistic data processing. In the final chapter Stevens attempts to give an overall appraisal of the state-of-the-art of automatic indexing. An excellent bibliography is appended.

The state-of-the-art survey is a useful and valid approach to the literature of a narrow subject area. This particular report is a good example. It fulfills its stated objectives within the inherent limitations of the approach and provides a well-organized and thorough overview of the field. It should be an excellent point of departure for the person who wants to explore the subject in greater depth. While the summaries do not make reading of the original literature unnecessary, they will guide the searcher to his appropriate place of interest. Coverage of the literature can be considered highly satisfactory if the fact is accepted that there is really no such thing as completeness in bibliographic work. There are 662 items listed. Perhaps the greatest strength of the report lies in its superior organization.

In summary, the survey is good and should be useful for those who work in the field as well as for those in related areas. We need more similar works to cover other topics in documentation.

Susan Artandi
Graduate School of Library Service
Rutgers, the State University

10/65—2R **The Future of the Research Library.** 1964. Verner W. Clapp. University of Illinois Press, Urbana. 114 pp. $4.50.

This 114-page book is deceptively small in relation to the many and complicated matters with which it deals. These all relate to the central problem of putting the research library of the future to greater use. The author, Verner W. Clapp, has been for the last nine years the president of a nonprofit organization, the Council on Library Resources, which is dedicated to making grants for "solutions to the problems of libraries generally and of research libraries in particular." For these purposes, the Ford Foundation has allocated thirteen million dollars to the Council. Thus, Mr. Clapp has had to live daily with decisions as to what, how, when, where, and by whom, work on improvements should and can proceed.

Mr. Clapp worked for 34 years at the Library of Congress, gaining experience in both workaday matters and policy decisions. He is a humanistically trained scholar who also likes to tinker. Thus, the author is a modern man, belonging to both of C. P. Snow's two cultures.

The Council and Mr. Clapp have concentrated on those improvements of library services made possible by new machines and techniques. It is always easier to talk about machinery than about where and how it can be usefully applied; and it is even harder to discuss what is needed in a particular field. The author does all three.

If anything, he is an optimist concerning machine possibilities. It is just because of such determined optimism that many things have ever been mechanized. An example of optimism about a little thing that would be wonderfully useful is the author's unending pursuit of an "adequate reading device for individual use" with high-reduction microtext.

An example of his optimism about a major problem is his enthusiasm for mechanized indexing:

> Suffice it to say that there are now experimental results which promise successful subject indexing of natural text by data processing machines at reasonable cost, providing that the text to be analyzed is available in machine-readable form.

The work to which he refers is that of Dr. D. R. Swanson. It would be interesting to learn what Mr. Clapp thinks of the less optimistic conclusions reached by Dr. John O'Connor in his work on mechanized indexing.

Clapp has two touchstones by which to guide himself and others through the tangles of new techniques and the competing claims for help. They are service to users and the cost of these services. He is obsessed, as any good librarian must be, with getting books into the hands of users at the cheapest possible cost. Though seldom calling the users by their older and apter name of readers, he well knows what the main business of libraries is.

He identifies two complementary principles governing library development: local self sufficiency and the sharing of resources. To him, the research library is typified by the university library, in which breadth of subject interests combine with the demands of systematic learning. A central and continuing interest of the author is to preserve library materials in their original form whenever possible, and then to supplement them with microcopies as needed.

Difficult solutions are neither ignored nor evaded. The author is not one of the fashionably guilt-ridden who insist on assuming the blame and the responsibility for everything. He is humble enough to realize that modern library problems are only a small part of the larger problems of modern technology and modern living:

The answer lies not only in the nature of scholarly research, but increasingly in the nature of all human activity, especially at the point at which mankind has now arrived: in his industrial and commercial activities; in the arrangements with which he governs his relationships to his fellows, locally, nationally, and internationally; in his relationships with the universe of nature upon which he is dependent. [p. 14]

For instance, he realizes that agreements to share resources among a neighboring group of libraries, or those at the national level (Chapter III), are complicated and protracted undertakings in the best of circumstances. But such agreements must be made, are being made and implemented, and even more are needed, as he repeatedly emphasizes. He also has the frankness to admit that some libraries operate quite effectively with patience and skill making up for gaps in budgets and distances between geographical locations.

Missing from the book is a very important and essential element. This is a subject index. For this little book with so much matter compressed in it, an index is a *must*. Therefore, the rest of this review is an effort to remedy this defect. The index is provided, too, by this superstitious documentalist, as a gesture toward undoing the curse placed on our infant field in the late 1940's—when an English librarian, Samuel Clement Bradford, wrote the first book entitled "Documentation" *without* a subject index.

LEA BOHNERT
C-E-I-R, Inc., and
The American University

10/65—3R **Some Fundamentals of Information Retrieval.** 1965. John R. Sharp. Andre Deutsch (A Grafton Book), London. 224 pp., index. $8.75.

The pleasant and systematic presentation made by Sharp makes this book seem even shorter than it is—a most engaging feat. The style is informed and literate, and the selection of references for each chapter is quite well made. There is a quite good (quite conventional) index, with relatively few undifferentiated page number entries.

The fundamentals Sharp discusses are limited almost entirely to methods of providing subject access points. His chapter heads are indicative of his drift: The Nature of the Problem; Classification; The Alphabetical Subject Catalogue; The Shortcomings of Conventional Systems; The Advantages of Non-conventional Systems; The Control of Indexing Language; Coordination and Correlation; Auto-Indexing and Auto-Abstracting; Conclusion.

Note the logical progression here. Note, too, the emphasis on concepts rather than devices. It is true enough, as it would be with any book on this subject at this time, that the reader (any reader) will find much to quarrel with in interpretation, in emphasis, in inclusion or exclusion of particular things or works. To this reviewer, for example, the book does not show the depth of understanding of alphabetical subject indexing which would be desirable—to others, the same might be true for faceted systems or vocabulary control.

The book is remarkable, however, for covering so much as well as it does. This is a more connected view than we have had before. If the lack of understanding of subject cataloging may be described as British, so may the occasional acceptance of the utility of particular technological wonders in the U.S.—the latter perhaps largely due to lack of opportunity to inspect; a situation which indicates that, wonderful as a knowledge of the literature may be and however important information retrieval is, it will never replace either baseball or the face-to-face encounter, be it with console or man.

Unlike many British works, this one is not devoted simply to spreading an English version of the Ranganathan gospel. Perhaps for this reason, both chain indexing and facet analysis, as well as other concepts of the Indian master, seem to come off better than usual, set properly in a perspective of other ideas of importance in the field.

The text provides copious, well-chosen examples. The author shows judicious personal skepticism. Perhaps the only significant omission of a type of indexing is that of the citation index, which is not mentioned at all—it would have been interesting to hear Sharp on this subject. Considerable consideration is given to the verbal form of headings, an area usually neglected. The identification of non-conventional with coordinate systems is not peculiar to Sharp, even though it is peculiar.

Overall, however, the book is a treat. It is full of piquant statements such as: "The difficulties which beset bibliographical classification cannot be alleviated by the application of non-conventional principles." (Which can, however, serve other purposes.) Or: "A classification scheme must not be modified to the slightest extent to suit the notation . . . any concessions . . . may mar the quality of the scheme."

It is a pleasure to see Fairthorne fairly represented, and Metcalfe acknowledged as an authority, all in one book with Vickery and Ranganathan and Shaw. It is also a pleasure to read the brief, piquant chapter on auto-indexing and abstracting, with its interesting conclusion. If too many examples come from (apparently) Cranfield or British Nylon Spinners, Ltd., so what. A book to own, consult, and enjoy. As ubersicht, ganz gut.

T. C. H.

ERRATA

"Measuring the Reliability of Subject Classification by Man and Machines," *American Documentation*, Vol. 15, No. 4, October 1964, p. 271: The formula for $r_{\text{♂}a}$ should read:

$$r_{\text{♂}a} = \frac{r_{am}}{r_{mm}}$$

"Random Code Numbers for Universal Identification of Documents," *American Documentation*, Vol. 16, No. 2, April 1965, p. 94: "The number of doubly and more highly assigned code numbers will be small only if $n \curlyvee u$ . . ." should read

The number. . . will be small only if $n \ll u$ . . .

INDEX

Index to Verner W. Clapp, *The Future of the Research Library*
(continued on next page)

Index to Verner W. Clapp, *The Future of the Research Library*
(concluded)

Brief Communications

Some Infract, Mire on Facts

Ralph R. Shaw's attempt to clarify the differences, if any, between information centers and special libraries, which he titled with the anagram "Face Its Norm,"[1] was, in the main, a thoughtfully composed paper—yet it somehow managed to miss the point almost completely. Also, it passes off as fact certain unsupported and highly unwarranted misinformation concerning at least one existing information center; hence, the two alternate anagrams offered in the title above.

In his paper, Professor Shaw lists six alleged characteristics of information centers derived from published and unpublished literature (unspecified) and from personal observation. He points out that none of these characteristics, nor any combination of them, is different from activities which could be found in special libraries in the past. He does find that information centers are distinguished from libraries in their use of people who exercise specialized judgment, of a type not required of a librarian, in analyzing materials which might be found in libraries. Seizing on this as the sole difference between information centers and special libraries, he implies that an information center is foolish or deficient if it does not always employ established library practices and trained librarians for all tasks other than the analysis of raw technical data.

Had he pondered the question of *why* information centers make use of specialized judgment, I am sure that Professor Shaw would be less sweeping in his indictment of unorthodoxy. He would also recognize, I think, that some of his alleged characteristics of information centers are stated incorrectly.

Information centers are usually established in recognition of some *existing, narrowly-defined, and urgent application* for technical information which is believed to be in existence but inadequately available. There is indeed a narrow area of specialization in most information centers as Professor Shaw indicates in his criterion 2; but the specialization is with respect to the *application* of data, and only incidentally with respect to the skills which are to be used in analyzing the data. Most information centers require specialists from a wide variety of disciplines. Putting it another way, the technical skills and procedures required in an information center are selected to serve the mission, or objective, of the center. The important thing is to get the job done by any means which seems appropriate, and not to limit oneself to the better known skills and methods and merely hope for the best.

Obviously it now makes sense for an information center to question the adequacy of traditional librarians and established library procedures *for its specific objective,* just as it makes sense to question the adequacy of skilled technical people and their specialties. When the objective of the organization is multifaceted or diffuse, when the level of pressure for immediate results is comparatively low, when the pertinent information is relatively easy to identify and acquire, the best course of action may well be to adopt standard library practices. On the other hand, when the objective is narrow, the pressure and economic stakes are high, and the pertinent information is hidden away in places which are unfamiliar to librarians, it is only common sense to turn over a large share of the work to the technical specialists, since they are the only people who can do it fast enough. These people may cost more in wages than their librarian counterparts, but this may be trivial compared to what is at stake in the larger situation which caused the establishment of the information center in the first place.

From the foregoing, it follows that Professor Shaw's criterion 5 is also incorrect. Information centers are not "designed and operated by scientists for scientists." An information center is unusally designed for engineers, physicians, or other applied technicians; and it is designed and operated by such people together with any scientists and information people who can carry their own weight.

On this last point, Professor Shaw has unfortunately made some inaccurate statements regarding the Titanium Information Center: the institutional memory of the larger Titanium Metallurgical Laboratory (TML) information clearinghouse project, for which I was responsible from its inception in 1955 until 1959 (by which time the project had been broadened to become the Defense Metals Information Center). He attributes to unidentified people and to unspecified reports the claim that we had "physicists and chemists performing order routines, cataloging, answering routine reference questions, and doing a number of other necessary tasks . . . which could been done by people trained for these jobs without waste of scarce scientific manpower." Since I do not know his sources, I cannot prove that such statements were never made; but it appears that the facts have been distorted virtually beyond recognition.

As a point of information, I quote here from a very explicit description of the Titanium Information Center which was published in a very well-known journal in 1956; and which still remains, I believe, an accurate description of Battelle's information centers.[2] The italics are added.

> ". . . Because it was Battelle's technical staff that was to make the most use of these files, the best way to insure meeting their needs was to have them work directly with the information specialists in developing the files . . . the engineers who use the information also constitute a part of the staff which puts information into the Center. This teamwork approach is the only assurance that information will be filed in the places where the users will look. . . .
>
> "A newly acquired item is given to an engineer who has specialized in the principal technical field covered by the article. *He thus has the opportunity* to use the data at once on current problems and to make notes for his private use. He *may* also indicate portions of the article that seem most significant before he passses the item to his teammate, the information specialist.
>
> "The information specialists, or analysts, who have been working closely with the engineers, have been made aware of the broad aspects and requirements of the project. They prepare the extracts, making the fullest use of notations made by the engineer. . . .
>
> ". . . the method is a unique one because it recognizes the concept of teamwork in a field where it has been lacking heretofore. A group of clerks, information specialists, and engineers has been developed into *a well-coordinated team* which makes certain that new information is integrated into the file and ensures economical and efficient retrieval of information and data on demand.
>
> "The information system . . . discussed here is not necessarily the one that would be required for another problem. However, *it is flexible*. . . ."

Professor Shaw's allegations imply that the technical staff was used for routine tasks instead of information specialists or librarians. This is not so. Professor Shaw's gracious concession that the situation has since been "largely corrected" is quite unnecessary. The situation never existed in the first place. Technical specialists and information specialists worked together in close cooperation. Tasks which could obviously be han-

dled better by one or the other were thus divided. There was an urgent job to do, and everyone involved worked hard and intelligently to get the job done in the most expedient manner. Talent was always welcome, whether technical or otherwise. A technical staff member (most of those on the project were metallurgists, not physicists and chemists) would not handle routine acquisitions work—that was done very well by the library. But the technical man might well make a trip to an engineering laboratory in order to coax, cajole, or horse-trade some undistributed reports out of their files. A librarian would have gotten nowhere. On the other hand, when an established library was reported to have a good collection of material on titanium, someone from our library would be expected to investigate. Technical staff members were expected to participate in the work of extracting and indexing; most of them, exercising common sense, limited their participation to advising information counterparts as to their special interests and requirements. The review of new documents by the technical staff, prior to extracting, expedited the dissemination of urgently needed information. And more than a few of the junior technical staff members achieved "expert" status largely as a result of the intensive reading they had to do for TML or DMIC.

Technical staff members did not answer all routine reference questions. Many were answered by the project's administrative staff, who had numerous review reports at their disposal. Others were routinely answered by the information specialists, especially when it was a request for information from specific documents. But the technical staff was generally given the opportunity to decide whether or not a reference question was indeed routine. In the early stages of the work particularly, a simple reference question from some remote laboratory or mill, when properly followed up by an engineer or scientist, might turn up the existence of an important titanium research program or titanium application which no one had known about.

When one reflects that TML was established as an extreme measure in order to help save a billion-dollar defense industry beset by more technical problems that it could cope with and on the verge of panic, the use of technical people in this manner is not wasteful of scarce manpower. It is the only way to assess the magnitiude of the problems quickly, to determine the availability of solutions, to disseminate the findings rapidly and effectively, and thus to stop the waste of scarce manpower on a colossal scale.

In summary, I offer the following quotation[3] from John R. Townsend, erstwhile Special Assistant to the Director of Defense Research and Engineering, whose comments are pertinent to both of my objections to Professor Shaw's paper (i.e., failure to recognize the importance of mission in the design of information centers, and misinformation on the appropriateness of procedures used at TML):

> ". . . There is a necessary balance among the various levels of research and engineering. And to bridge the gap from one level to another requires that we give a great deal of attention to the means of communicating among them. . . .
>
> "For the scientist, the information need centers on completeness of coverage, and reasonably adequate retrieval of desired publications. But for the engineer the information needs not only to be assembled; it must also be evaluated, collated, put in usable form, and placed in his hands, all without delay. . . .
>
> "With these thoughts in mind, we undertook an experiment. . . . We established . . . a Titanium Metallurgical Laboratory charged with serving all persons and companies whose needs for technical data on titanium were directly or indirectly connected with defense. . . . [It was assigned] the functions: (a) of collecting all useful data, published and unpublished, concerning titanium and possible competitors; (b) of visiting persons and companies engaged in work on titanium to exchange data . . .; (c) to do . . . laboratory research of . . . reconnaissance nature . . . in support of the information function; (d) to answer responsible questions . . .; (e) to supply . . . information on the various engineering phases of titanium; (f) to issue occasional papers of somewhat more speculative . . . nature . . . directing engineering attention to important but incompletely solved phases of titanium development. . . .
>
> "The device proved so valuable to the Department of Defense that . . . we expanded the activity . . . into the Defense Metals Information Center. . . . In this expanded activity, the Center is proving of even greater value. . . ."

BEN-AMI LIPETZ
Consultant, Carlisle, Mass.

References

1. R. R. Shaw, "Face Its Norm," *American Documentation,* Vol. 16, 1965, pp. 77-80.
2. R. W. Gibson, Jr., and B.-A. Lipetz, "New Look in Manual Methods," *Special Libraries,* Vol. 47, 1956, pp. 106-113.
3. John R. Townsend, "Technical Information and the National Defense," Address before the Science-Technology Division, Special Libraries Association, June 8, 1960.

References

1. G. SALTON, "A Document Retrieval System for Man-Machine Interaction," *Proceedings of the 19th ACM Annual Conference,* Philadelphia, 1964.
2. G. SALTON, and M. E. LESK, "The SMART Automatic Document Retrieval System—An Illustration," *Communications of the ACM,* Vol. 8, No. 6, June 1964.
3. G. SALTON, et al, *Information Storage and Retrieval,* Reports No. ISR-7 and ISR-8 to the National Science Foundation, the Computation Laboratory of Harvard University, June 1964 and December 1964.
4. J. A. SWETS, *Information Retrieval Systems Science,* Vol. 141, No. 3577, July 19, 1963.
5. C. W. CLEVERDON, "The Testing of Index Language Devices," *ASLIB Proceedings,* Vol. 15, No. 4, April 1963.
6. ———, "The Testing and Evaluation of the Operating Efficiency of the Intellectual Stages of Information Retrieval Systems," International Study Conference on Classification Research, Elsinore, September 1964.
7. D. R. SWANSON, "Searching Natural Text by Computer," *Science,* Vol. 132, No. 3434, October 21, 1960.
8. D. R. SWANSON, "The Evidence Underlying the Cranfield Results," *The Library Quarterly,* Vol. 35, No. 1, January 1965.
9. J. ROCCHIO, "Performance Indices for Document Retrieval Systems," *Information Storage and Retrieval,* Report No. ISR-8 to the National Science Foundation, the Computation Laboratory of Harvard University, December 1964.
10. R. A. FAIRTHORNE, "Basic Parameters of Retrieval Tests," short paper presented at the ADI Annual Meeting, Philadelphia, October 1964.
11. J. A. SCHULLER, "Experience with Indexing and Retrieving by UDC and Uniterms," *ASLIB Proceedings,* Vol. 12, No. 11, November 1960.

Letters to the Editor

Dear Sir:

James van Luik's comments ["The Chemical Abstracts KWIC Index," *American Documentation*, Vol. 16, No. 2, April 1965, pp. 122-23] on Chemical Abstracts Service publications indicate some lack of understanding of the objectives of the American Chemical Society's Chemical Abstracts Service (CAS). Without addressing these remarks to the objectives of each of the individual publications and services, it may be stated that the overall objective is to provide a full range of chemical information services to the scientific community. The individual information packages are interrelated; each is aimed at a specific need, but it is so constructed as to allow independent use.

Chemical Abstracts itself, published biweekly, classified into 74 (not 70) sections, and available either as a whole or in 5 individual groups of sections, serves as both a current awareness and a retrospective searching tool for all of chemistry (some 10,000 serial publications plus patents from 24 countries are covered). To enhance its usefulness in alerting and to supply interim search capability until the volume indexes become available, the biweekly issues have included for the past two-and-a-half years a man-produced keyword [not a computer-produced keyword-in-context (KWIC)] subject index with index terms derived from the text and the title of the abstracts. In order to assure currency (median abstracting time for all of *CA* is now 105 days), no attempt is made in the keyword index to produce a sophisticated, in-depth tool; the volume indexes are intended to provide the permanent means of access to *CA*. However, the biweekly keyword indexes, used in conjunction with the *CA* cross-reference system and the classification of issues into 74 specialized sections, provide the user with prompt access to that material which he needs to support his work, although at the expense of more of his personal effort than will be required in using the volume indexes.

Reading the titles in a given section provides broad access to the papers related to the subject area specified in the title of the section, but the arrangement into sections follows a hierarchy which has all the limitations of any hierarchy. The cross-reference system and the issue keyword index allow the *CA* user to break the pattern of the established arrangement into sections. On the other hand, to drop the abstracts as suggested would render a disservice to the scientific community at large. Abstracts, especially informative ones such as *CA* publishes, are designed to provide an adequate basis for appraisal of the technical contents of the original papers and thus to simplify and to hasten the selection of those papers which deserve detailed study.

Chemical Titles (*CT*) is not intended as a replacement for *CA*, but rather as an accessory aimed at increasing the speed with which an investigator can get to the chemical contents of approximately 650 productive scientific serials. While *CT* serves an alerting function, its indexes are neither deep enough nor well enough organized to provide permanent access to the chemical papers which are covered.

The volume and collective indexes to *CA* are intended to provide the required permanent, efficient route of access to the whole of the chemical literature, namely the contents of *CA*. A substantial part of *CA*'s value is associated with its complete, unbiased, consistent coverage of the chemical literature. Without these characteristics the scientific public would soon lose its faith in the product—this view has been emphasized by every survey of *CA* users. The current CAS program to convert *CA* to a computer base is aimed at increasing the usefulness of *CA* and its indexes by increasing the accessibility to the included information by decreasing the rigidity of present hierarchies.

Finally, there is always the implication in comments such as Mr. van Luik's that the reference information service is at fault because the volume of information is great. No one really thinks this, of course, but the implication is there. What is really needed as the volume of research information continues to expand is a thorough, careful, but non-critical agency to collect, bolt-down and categorize what the scientists are producing. The scientists, then, can select—find and choose to use or ignore what is presented to them. This is an abstract of the primary function of CAS.

GERALD SALTON
The Computation Laboratory of Harvard University

June 8, 1965

Dear Sir:

On behalf of (B')'S (the Not-Not-Boolean Society), I would like to present an antithesis to Prof. Donald J. Hillman's thesis (*American Documentation*, Vol. 15, No. 3, pp. 217-225) to the effect that Boolean Algebra is inappropriate as a model for information systems. I have pondered Hillman's views (op. cit. and his Reports 1, 2, and 3 cited in Reference 4 of op. cit.) and those of Fairthorne and Mooers. (R. A. Fairthorne, *Towards Information Retrieval*, London: Butterworth, 1961.) My admittedly partial coverage of the literature leads me to suppose that there may be a concensus in favor of abandoning the use of Boolean Algebra for information system models. I can't help wondering, however, if my predicament is unique; perhaps I am not the only (B')'S-er who feels that the obsequies for Boolean models are premature.

To simplify matters, this discussion will assume a Boolean algebra, as formulated by Langer (S. K. Langer, *Symbolic Logic*, Boston: Houghton Mifflin, 1937, Chap. IX, esp. p. 233ff.) applied to an information system that shall be left unspecified except that, (a) the items whose storage and retrieval are the system's reason for being shall consist of information embodied in discrete documents, and (b) the black box shall include an index consisting, essentially, of an array of criteria, or indices, which, in Boolean combination, shall specify the end in view of all searches that can validly be made through the system.

Complementation, the *fundamentum divisionis*, the Law of Excluded Middle, the dichotomy between A and not-A, the two-valued logic (Langer, op. cit., p. 142ff.): these are various ways of indicating the target of the anti-, or post-,Booleans. What they say apparently amounts to this: (1) You can't complement an information system index class A because, in information systems, there can be no determinate universe class I that can be dichotomized into A and its complement not-A (or A') (Fairthorne, op. cit., pp. 125ff.); and (2) if complementation is not defined no Boolean model is applicable.

The validity of the second of the above assertions I grant on purely formal grounds. With respect to the indeterminacy of the universe class in information systems, I propose to define membership in the universe class (i.e. the "intensional" class: Langer, op. cit., p. 130) in such a way that the actual membership (i.e. the "extensional" class: Langer, ibid.) shall be fully determined for all purposes appropriate to an information system, as follows:

> A necessary condition that a document, d, shall be a member of the universe class, I, of an information system, S, shall be that d has been fully indexed according to the prescribed procedure of S before being entered in the store of S.

Fairthorne (ibid.) seems to feel that any such condition on membership in I is inadmissable; I cannot for the life of me see the cogency of this objection, though I assume it is based on vast experience (i.e., let's have chapter and verse). Without going

into a general discussion of the trade-offs that are necessarily involved in system design, I submit that the restriction of system membership to indexed items is not very onerous; the line between what is and what is not in the system must be drawn somewhere. Only confusion can result from including partly processed and not-yet-processed documents in the system. The very system concept would seem to preclude any such inclusion.

System membership has been discussed above without explicit regard for its dynamics. Let I_0 be the extension of the universe class of a system at the instant when a search, s_0, is initiated. Let I_1 be the extension of I at the instant when s_1 is begun. Whether I_0 and I_1 are identical, or either one is larger, depends on whether any input or discarding takes place in the time interval from instants 0 to 1. Input and discarding may be virtually continuous or discrete, regular or irregular; in any event, while s_n is in progress, I remains in its nth state, I_n. Whether the temporal succession of states, $I_0, I_1, \ldots, I_n, \ldots$, is conceived of as a discrete or a continuous succession will depend on various characteristics of the system, including relative continuity of input, typical and maximum search times, etc. I am inclined to favor the somewhat paradoxical concept of the successive states of I being both finite in temporal length and, virtually, instantaneous with respect to the continuity of I over time. This is, perhaps, a romantic view based on inadequate acquaintance with retrieval hardware. Input is so infrequent and spasmodic in the systems with which I am familiar (in the National Archives of the U.S.) that the conception of discrete states of I does no violence to the reality.

Given that I_n is determined, let A be the extension of a search prescription in I_n. Then A', the complement of A in I, is determined, and Boolean information system models are therefore admissable. Whether or not they are expedient choices as models depends on questions of purpose and trade-off that will not be gone into here. Presently available alternatives to Boolean models are not very appealing; when new models come along they must outperform, before they can supersede, the Boolean models. The (B')'S rests its case and adjourns *sine die*.

Yours truly,

C. E. Dewing
McLean, Va.

P.S. On looking over this letter I am a little overwhelmed, but not sufficiently to make me refrain from sending it to you. I have no serious hope that you will want to print any of it. Please regard it as a *cri de coeur* from one of your most unsophisticated readers. No reply is necessary. My last word is: Couldn't you prevail upon Fairthorne to elucidate, with empirical evidence, his reasons for disliking Boolean models? Does such an explanation exist? I have not found it.

C.E.D.

May 28, 1965

Dear Sir:

The "Note on the Pseudo-Mathematics of Relevance" by Mortimer Taube, published in the April 1965 issue of *American Documentation* was originally distributed at the NSF study conference on Evaluation of Document Identification Systems and Procedures, held in Washington during October 1964.

It was pointed out to Dr. Taube at the time that his whole case rested on a confusion between two distinct uses of the term "relevance." In an information retrieval context, these two types of relevance may be described as follows:

relevance 1 (hereafter termed "relevance") involves a judgment made by a user of an information system when he decides that a given document retrieved in answer to a search request is either useful to him (relevant) or not useful (irrelevant);

relevance 2 (hereafter termed "precision") is a measure proposed for the evaluation of the retrieval effectiveness of a system, and is defined as the proportion of retrieved documents actually relevant to a user's purpose.

Now it is perfectly true that precision (relevance 2) depends on relevance (relevance 1), since it is impossible to ascertain the number of documents both retrieved and relevant, unless someone has first determined which documents are relevant to a given request and which are not. It is, however, equally true that once this relevance judgment is made, say by some given user A (once relevance 1 has been determined), then precision (relevance 2) is a perfectly unambiguous measurement, applicable to determine the effectiveness of the system from the point of view of user A. Moreover, if user A is accepted as a typical user, then the precision measurement is, in fact, a valuable indicator of system effectiveness for other users as well.

If, on the other hand, user A is not accepted as typical, then precision may still be determined unambiguously, by having relevance assessments made by a test group of users, and by labeling a given document as relevant to a certain request if a majority of the test group agrees that it is. These relevance judgments by consensus can then be used as before to compute a useful precision value.

Failure to distinguish between relevance 1 and relevance 2 leads Dr. Taube into a totally uninterpretable discussion, in which he talks about an "illegal shift from subjective relevance to mathematical relevance." If the basic definitions are understood, there can be no question of a "shift," legal or illegal, and the recall and percision measures must be accepted as what they are: perfectly reasonable indicators of system performance.

It is likely that the Cranfield studies and the A. D. Little work can be attacked on technical grounds, for many good and solid reasons. A diatribe, using Webster as an aid, and based on misinterpretation or misunderstanding of the fundamental definitions does not appear to this writer to be the best way of discrediting other people's efforts.

Sincerely,

Gerard Salton
Computation Laboratory of Harvard University

July 12, 1965

Dear Sir:

In the second paragraph of Dr. Salton's letter, he says: "It was pointed out to Dr. Taube. . . ." The "it" should be changed to "I." If the "it" is allowed to stand, it makes it appear that the study conference came to this conclusion rather than Dr. Salton individually, and this is certainly not true.

Although I don't think the distinction between relevance 1 and relevance 2 is meaningful in an operating situation, as opposed to an experimental situation in which all documents are tagged in advance as being relevant or irrelevant, Dr. Salton is here agreeing with me that the Cleverdon and Little papers used two meanings of the term "relevance." Since the actual papers did not explicitly point out the use of two distinct meanings, I fail to see how I can be accused of confusing what was never distinguished. It might be an interesting exercise for Dr. Salton, or anyone else, to go through the Cleverdon Reports and the Arthur D. Little Study and mark each use of the terms "relevance" and "recall" with "1" or "2." Such an exercise would do more to demonstrate the invalidity of these papers than anything I have said in my note in *American Documentation.*

Faithfully yours,

Mortimer Taube
Documentation Incorporated

LITERATURE NOTES

BURTON E. LAMKIN, Editor

Contributors of abstracts from readers and suggestion of books and articles for review or inclusion in this bibliography will be welcome and are actively solicited. Volunteer abstractors and reviewers are needed, as well as people with special linguistic talents. This issue contains abstracts submitted to Literature Notes by the Division of Chemical Literature of the American Chemical Society. All copies of reprints, reports, and correspondence for this section should be addressed to Mr. Burton E. Lamkin, 1649 Fairorchard Ave., San Jose, Calif., 95125. In order to increase its coverage, this section will include some abstracts copied or adapted from other publications. The American Documentation Institute is not able to supply copies of the publications abstracted or cited.

Abstractors

The following persons have prepared some of the abstracts for this issue. Their names are listed here to express the editor's thanks and appreciation, as well as to explain the initials that were used to sign the abstracts.

S.A.	Susan Artandi	C.B.L.	Charles B. Leidenfrost
C.L.B.	Curtis L. Brown	M.C.P.	M. Constance Parché
T.G.F.	Trudy G. Fisher	F.S.	Frantz Spachner
K.G.	Kurt Gingold	D.C.W.	David C. Weber
M.H.G.	Maria H. Grimes	H.L.W.	H. L. Williams
A.B.J.	A. B. Johnson		

Contents

Starting with the January 1965 issue, separate copies of this section of *American Documentation* may be obtained as reprints on a subscription basis for $4.00 per year. Requests should be directed to ADI, 2000 P Street N.W., Washington, D. C., 20036. It is hoped that in the very near future it will be possible to provide Literature Notes as a publication separate from *American Documentation* and with increased coverage. The availability of separate reprint sections is a start toward that objective.

Index of Authors and Corporate Sources Referred to in this Issue

(The numbers refer to the terminal digits of the abstract serial numbers in this issue. The letter R refers to Reviews.)

Journal Abbreviations

Many of the journal titles are represented by their corresponding CODEN notation as defined in the two publications, *Coden for Periodical Titles,* American Society for Testing and Materials, Phila., Pa., 1963, ASTM special technical Publication No. 329, and *Coden for Periodical Titles, Supplement 1,* 1964, ASTM Special Technical Publication No. 329—31. The CODENS are being used in place of the full titles primarily to reduce typing and typesetting work, and to provide a more compact representation of the citation. The following table will help translate the CODENS for this issue.

ALBL	ALA Bulletin
AMDO	American Documentation
ARHE	Arthritis and Rheumatism
ASLP	Aslib Proceedings
BABL	Babel
BATR	Battelle Technical Review
BEHS	Behavioral Sciences
BHOR	Business Horizons
BLRC	Bell Laboratories Record
BMLA	Bull. Medical Lib. Assoc.
BRNS	Barron's
BUAT	Business Automation
CACM	ACM Communications
CENE	Chemical and Engineering News
CENG	Control Engineering
CMPD	Computer Design
CMPG	Computer Journal
CPGR	ACM Computing Reviews
CRLI	College and Research Libraries
DKMN	Dokumentation
DTMN	Datamation
ERGO	Ergonomics
FEPR	Federation Proceedings
FRON	Frontier
HABR	Harvard Business Review
IBMJ	IBM Journal
IEAI	IEEE Trans. on Applications & Industry
IEEI	IEEE Trans. on Ind. Electr. & Control Instr.
IEEP	Proc. IEEE
IETT	IEEE Trans. on Information Theory
IEEW	IEEE Trans. on Eng. Writing & Speech
IFCN	Information and Control
IPAL	Inland Printer and American Lithographer
IREL	IEEE Trans. on Electronic Computers
IRTE	IEEE Trans. on Eng. Management
ISCT	Int'l Science & Technology
ISSR	Information Storage & Retrieval
JACM	ACM Journal
JCHD	J. Chemical Documentation
JDOC	J. Documentation
LANS	Language and Speech
LCIB	Library of Congress Information Bull.
LIBJ	Library Journal
LIBL	Library Literature
LIBQ	Library Quarterly
LIBT	Library Trends
LRTS	Library Resources & Technical Services
MEEN	Mechanical Engineering
MTRN	Mechanical Translation
MTRS	Materials Research & Standards
NADO	Nachrichten fur Dokumentation
NAMN	National Micro-News
NATU	Nature
NATW	Naturwissenchaften
NMRV	Navy Management Review
NWSC	New Scientist
OPRE	Operations Research
PMOS	Perceptual and Motor Skills
PRMG	Printing Magazine/National Lithogr.
PSEN	Photographic Science and Engineering
REDE	Research/Development
REDO	Revue Internationale de la Documentation

REPM	Reproduction Methods for Business & Ind.	SHBN	Stechert-Hafner Book News	TAPP	Tappi
RESM	Research Management	SPLB	Special Libraries	UNDL	UNESCO Bull. for Libraries
RPRV	Reproduction Review	SSTM	Systems	WLDU	Wilson Library Bull.
SCIE	Science	STNW	Sci-Tech News		
		TAGU	Trans. of the Amer. Geophysical Union		

Conference Proceedings and Collections of Papers

10/65–1 **Report on the 16th Annual Convention and Labor Conference of the German Society for Documentation,** e.v. at Bad Durkheim, Weinstrasse, Casino, from October 5 to 8, 1964.

The general theme of the convention was "Documentation and Its Users." (*In German.*) Bericht uber die 16. Jahresversammlung und Arbeitstagung der Deutschen Gesellschaft fur Dokumentation e.V. in Bad Durkheim an der Weinstrasse, Kurhaus, vom 5. bis 8. October 1964.—Helene Eckard, NADO, v. 15, no. 4, Dec. 1964, p. 192-200. (BATR)

10/65–2 **Proceedings of the 1963 Clinic on Library Applications of Data Processing,** Illini Union, Urbana Campus, University of Illinois, April 28-May 1, 1963, Herbert Goldhor, ed., distributed by the Illini Union Bookstore. Champaign, Ill.

10/65–3 **Seminar on Information and Documentation,** Gorisch, R., *Mitteilungen aus dem Wissenschaftlichen Bibliothekswesen der Deutschen Demokratischen Republik* (Reports on Scientific Libraries of the German Democratic Republic), No. 7, 1964, p. 80.

The Institute of Economic Sciences of the German Academy in Berlin held in June 1964 a seminar on the subject of information and documentation in the field of social sciences. Emphasis was placed on the various types of information activities, the role of documentation services and methods of setting up such a service, problems of classification, and the necessity of better information sources. Cooperation between libraries and documentation centers was stressed. An exhibition on information and documentation in the social sciences was organized. (DOC. INC.)

10/65-4 **UNESCO Meeting on Scientific Translation and Terminology.** R. W. Jumpelt. BABL 10(1):25-28 1964.

Describes the activities of a UNESCO working party convened in Rome January 27-February 1, 1964. Their discussions included a proposal for the international coordination of terminological activities. The project aims to arrive at agreed policies for improving lexicographical coverage and controlling the growth of terminologies in science and technology. The working party also reviewed the scope, nature, organization, and functions of existing regional, national, and international translation centers. (STNW)

10/65–5 **First International Meeting of Experts from Socialist Countries on Designing and Building of Libraries and on Library Equipment,** Gawrecki, D., *Technicka Knihovna* (Technical Library), No. 4, 1964, pp. 105-117. In Czech.

The author examines the cooperation and division of labor among all the socialist countries as proposed at the meeting in Prague on January 11-14, 1964. He suggests the following actions: preparing uniform data sheets on libraries in each country; listing essential materials and services available; standardizing technical equipment, amount of space, etc. which will aid planning and building; increasing the qualifications of experts through specialized courses in graduate schools; organizing visits to countries with the best developed library systems; and on-the-job training of teachers for specialized schools. Division of labor for theoretical, developmental, and research projects should be reviewed to permit more effective participation of experts of different nationalities. Serialized production of equipment and hardware is desirable. Design, planning, and production should be carried out jointly by several countries according to the field of specialization of each and under the supervision of an international coordination center. Information and documentation should be provided for by individual countries via national institutions. A handbook on designing and building libraries and library equipment should be prepared. For acquisitioning and exchanging information on library activities in individual countries all data should be centrally collected and the results published in a uniform manner. (M.H.G.)

10/65–6 **Communication Problems in Biomedical Research: Introduction and Report.** (Sept.-Oct. 1964) R. K. Cannan, FEPR 23(5): pp. 1117-32 Pt. 1.

10/65–7 **Conference of the Polish Academy of Sciences Libraries on the Management of Book Holdings,** Kelles-Krauz, J. and Zawialska, M., *Przeglad Biblioteczny* (Library Review), No. ½, 1964, pp. 131-135. In Polish.

The proceedings of the conference in Wroclaw, Poland, in May 1963, concerning the disposition and exchange of surplus copies of books and publications in and among the libraries of the Polish Academy of Sciences, are described. It was suggested that libraries retain three copies of a publication and hand the rest over to a central surplus store yet to be established. The central surplus store would redistribute the publications among libraries that have demand for them. Furthermore, it was proposed to establish a coordination center for local and international exchange of books and periodicals which could draw on the above store. Legal and practical aspects of handling surplus holdings were investigated. (DOC. INC.)

10/65–8 **131st Meeting of AAAS—Report of Section T.** (Montreal, December 1964) Phyllis V. Parkins and Gunther Marx. *SCIE 147,* 943-4, 946, 949-51 (1965). Reports on the sessions of Section T (Information and Communication) of AAAS. (*K.G.*).

10/65–9 **Foreign Libraries:** SPLB, *56*(2), 93-106 (Feb. 1965) A collection of papers including: *Guidelines for Consultants to Libraries in Developing Countries,* J. A. Aufdenkamp, 93-97; the Biblioteca *Mario Negri, Report from Milan* (in pharmacology and related fields), G. M. Clark, 98-102; Four Scientific Information Centres in Europe. (The European Translation Centre in Netherlands, International Documentation Centre in Sweden, National Lending Library and Oxford University School of Forestry in England), E. A. Keeley, 103-106. (*M.C.P.*)

10/65–10 **Second International Study Conference on Classification Research: Conclusions and Recommendations.** Winter 1965. *LRTS* 9(1):113–117.

Individuals from 16 countries and 2 international bodies met at Elsinore, Denmark, Sept. 14-18, 1964, to consider progress in the design of classification systems and in the application of machines to information retrieval. Important contributions were found to have been made to classification theory by such disciplines as structural linguistics, semantics, mathematics, logic, and epistomology. In the field of theoretical research, seven areas are identified for further investigation and two needing investigation are seen at the "frontier" of theoretical research and practical application. In the evaluation of classification systems, seven areas are identified, including three requiring international cooperation in a test of large collections. In the field of automated classification, statistically reliable studies are needed to determine scope and limitations in comparison with other classification methods. Seven recommendations are made. One of these indicates: "Shortage of personnel for classification research is a serious obstacle to progress. Measures to improve this situation by training are being made and should be encouraged in institutions of higher learning. Attempts should also be made to interest research workers from certain other disciplines in classification research problems." (D.C.W.)

10/65–11 **First Congress on the Information System Sciences. Session 10. Joint Man-Computer Languages.** March 1964, 169 p. Rept. no. SS 10. Contract AF33 600 39852, Proj. 704 ESD TDR63

474 10. Unclassified report AD-435 102, Div. 32, 30, OTS price $12.00.

10/65–12 **National Conference of the Scientific-Technical and Economic Information Centers. H. K.,** *Aktualne Problemy Informacji i Dokumentacji (Current Problems of Information and Documentation)*, No. 1, 1964, pp. 22–24. In Polish.

The conference took place in Warsaw, Poland, in December 1963, and was attended by approximately 300 persons. The main subjects of discussion were the future development and tasks of the technical and economical information network, the role of the trade unions in the information services, research activities on various levels of the information system network, and the activities of the information centers as they affect the national economy. A number of suggestions on how to increase the effectiveness of the information service were made. These include the necessity of a better coverage of foreign patents in the *Bulletin of the Most Important Foreign Patents*, and better information on the available information services in the country. There were complaints that the managers of the industrial enterprises do not appreciate the value and importance of the work of the information departments under their control. (DOC. INC.)

10/65–13 **Japan Steps Up Output of Scientific Journals.** Number of periodicals has increased to 2337, more than 70% of which have started publication since World War II. Describes results of U. S.-Japan Conference on Editing of Primary Journals held in March 1965 in Tokyo. CENE v. 43, no. 16, Apr. 19, 1965, p. 94-95. (BATR)

Documentation–General

10/65–14 **Manpower for Technical Information Work; a Pilot Study.** U.S. Dept. of Labor. Manpower Administration. Office of Manpower, Automation and Training. December 1964, 23 p. (Manpower/Automation Research Monograph No. 1.)

Presents the findings and conclusions of a pilot study of the functions and qualifications of technical information support personnel in the communications equipment industry. Study was performed for U. S. Dept. of Labor under contract no. OAM 7-63 by Auerbach Corp. of Phila., Pa. Copies of the research monograph are available from U. S. Dept. of Labor, Manpower Administration, Washington, D. C. or in Dept. of Labor field offices. The full report, "A Study of Manpower Requirements for Technical Information Support Personnel," prepared by the Auerbach Corp. for the Dept. of Labor in Jan. 1964, may be consulted at selected depository libraries or in Dept. of Labor field offices.

10/65–15 **Toward the Library of the 21st Century. A report on progress made in a program or research sponsored by The Council on Library Resources.** Bolt Beranek and Newman Inc., Cambridge, Mass., March 1964.

We do not know what the library of the future will look like or how it will work. We do know, however, that the world's store of literature now amounts to hundreds of millions of different books, and in the next century will reach billions.

We also know that sheer numbers cannot measure the entire burden of seeking and using information needed in the carrying out of intellectual pursuits, such as those in research, especially when those pursuits draw upon knowledge that obsolesces almost as fast as it is gained.

Further, we know that certain advances already made, or now envisioned, in science hold promise of alleviating burdens of information usage in all activities that draw upon resources of the library.

During the past two years, under sponsorship by the Council on Library Resources, a group of scientists at Bolt Beranek and Newman Inc. conducted a program of research addressed to these questions:

How can we formulate the tasks the library will face in the twenty-first century? What can the science of information handling do toward helping the library of the future perform its tasks in more effective ways?

This report describes progress made toward answering the questions.

The reader will find here neither a comprehensive review of related research going on throughout the world, nor a detailed presentation of the work performed in this program. References listed at the back of the report indicate some of the relevant literature and cite the technical publications that resulted from this program of research.

The report sets forth the main results of the research in general terms and interprets them within the context of library functions. Toward this end, a schematic diagram, developed on the next few pages, appears at several points throughout the report to identify each research topic mentioned with the particular function to which it applies.

10/65–16 **Broadhurst, P. L. Swimming Against the Flood of Literature.** *Bulletin of the British Psychological Society*, vol. 17, no. 56, July 1964, pp. 1-7.

The coverage of all the literature of a given science by one abstract journal and the formation of large computer-searched central stores of information are ideas that are unlikely to be implemented on a large scale in our lifetime. Measures to relieve the present situation are divided into public palliatives—current contents lists, KWIC indexing, citation indexing, and private palliatives of which post co-ordinate indexing is said to be the most useful development. The ways in which journal editors could help include: providing a standard reference card for each article as an integral part of each issue; giving complete addresses of authors on the contents page; citing both initial and technical page numbers on the contents page; dividing the subject matter of the articles contained in the issue by sectional headings; printing brief abstracts of papers with the title on the contents page. (JDOC)

10/65-17 Colinese, P. E. **The Information Scientist.** *Scientific Business*, vol. 2, no. 5, May 1964, pp. 30-8.

Two main topics are discussed: namely, the increase in communication problems which has accompanied the "information explosion" and emphasized the need for information scientists; and the work of the information scientist.

Communication problems arise from: the greatly increased quantity of information being generated; the greatly increased quantity of information being sought; the increasing complexity of lines of communication; the speed with which information is wanted. Additional complications have arisen because of the following factors: the surge of applications to protect discoveries by patenting; the tendency for large industrial research organizations to be set up where there is a lack of communication between the various laboratories within an organization; the need in large research organizations for scientists and engineers to explain and interpret their work so that management decisions may be made. The large number of original papers being submitted to periodicals has resulted in a corresponding increase in secondary publications. The knowledge and qualities that the information scientists should possess to cope with this state of affairs are considered. The work of the information scientists may consist of any or all of the following: search for information within and outside their own organizations in response to specific requests or in anticipation of users' needs: sift, evaluate, and select information; undertake various technical writing tasks (e.g. prepare literature reviews, abstracts, summaries, regular progress reports, and information bulletins); technical editing; make translations; organize the material they use which entails knowledge of classification and indexing schemes and of equipment which might aid them. (JDOC)

10/65-18 **A System for Retrieving Untitled Business Documents.** Louise Schultz. *Parameters of Information Science.* p. 453-456. (see 4/65-1)

Computer-based systems for retrieval of scientific documents and information presume a large data base and/or high input rate. To provide top management with a tool for timely, absolutely reliable access to items in a small but critical collection of untitled and dissimilar documents, a machine-oriented system called "co-occurence indexing" supports management of a major project at System Development Corporation. Co-occurrence in-

dexing of an hypothetical collection of management letters and memoranda is described in this paper.

10/65-19 **CAS to Study Registry of Chemical Compounds.** CENE 43, pp. 23-24 (June 7, 1965).

Under a two-year, $2 million contract with the National Science Foundation, the ACS's Chemical Abstract Service, under Director Dale Baker, will attempt to prove the technical feasibility of a computer-based "registry of chemical compounds." This registry system is to be the foundation of an over-all national chemical information network. The CAS registry is based on chemical structure. Its purpose is to be able to determine as, supposedly, new compounds showup if anything has been published on them before and, if so, what. (H.L.W.)

10/65-20 **Identifying Key Contributions to Information Science.** Cuadra, Carlos A. AMDO, vol. 15, no. 4, October 1904, pp. 289-95.

Several alternative approaches are examined. Advice from experts on who were the most important contributors may give useful short cuts if there is some means of selecting the experts and of taking their differing value systems into account. Current textbooks are also helpful, although they show considerable variation in point of view, subject coverage, and in the selection of references. Of 911 distinct references appearing in six current textbooks, only 14 per cent were found in more than one of the books. Various hypotheses are put forward to account for this, including one that writers repeat themselves often enough for a given idea to be used in several of their papers. The analysis of bibligraphies is another approach. Four of the most comprehensive bibliographies were examined. Authors which appeared at least twice in three of the bibliographies and at least three times in the fourth bibliography, which had many more entries, were selected for a frequency count. This resulted in 322 authors being included. The twenty-five authors who appeared most often in the bibliographies were listed by "publication density" score, which is the *product* of the four numbers representing the number of times references to these authors appeared in each of the bibliographies. The *product* rather than the sum was chosen as this gives greater weight to authors who are uniformly listed in the four bibliographies. A list of basic readings can be compiled consisting of the fifty-five publications of these authors which appeared in all four bibliographies. This list would omit material published outside the period covered by all four bibliographies and important contributions by the less prolific writers. The unclear selective patterns of the bibliographies also present problems. Citation indexes provides leads to publications, but they necessarily stress older publications. Another possibility is to ask experts to identify the actual key contributions, but this presents the difficulty of asking people to judge their colleagues' work on top of the difficulty of selecting the experts. The approach of asking the authors themselves to indicate which of their contributions they consider to be the most important is being investigated at present. Approximately forty authors with the highest publication densities have been selected for this purpose. One of the primary purposes of this study was to determine whether there was a common agreement on the conceptual, methodological, or practical contributions to information science. Although no conclusion can be reached from the limited evidence presented, the study suggests that this is far from being the case. (see 4/64-9) (JDOC)

10/65-21 **What Do You Mean I Can't Write?** John Fielden, HABR. 42, (3): 144-148, 151-152, 154, 156 (May/June 1964).

Reasons for ineffective writing at the executive level are discussed under the following headings: readibility (reader level, jargon, direction, focus, style); correctness (grammar, format, coherence); appropriateness (tact, diplomacy, opinion, attitude, motivation, etc.); and thought (preparation, competence, fidelity, analysis, and persuasion). CLB

10/65-21 **Mooers torvenye: megjegyzesek az informacio hasznalatanak lelektanahoz es szociologiajahoz** (The Moores-law: Contributions to the Psychology and Sociology in the Application of information), Pal Gyore, *Tudomanyos es muszaki tajekoztatas,* 11 (9-10) November-December 1964, 748-762. References.

Not much attention is being paid to the problem of how information is being used. According to the law discovered by Calvin Mooers, an information retrieval system will not be used until the user feels more uncomfortable and embarrassed while being informed than when not. To accept information requires thinking, decision, activity, effort. There are deep psycological reasons hidden behind this law. Everyone would like to maintain his peace of mind, but at the same time wishing something new would happen to him. Information, if novel and surprising, will excite, cause problems, and exert disturbing effects. Creative personalities will not be frightened by this feeling; those performing routine activities may also prove sensitive towards such phenomena. The latter type of person must be treated patiently with information supplied in adequate proportions to achieve sufficient influence. Creative ability and the capacity of information intake are proportional, thus the development of creative ability will improve the sensitivity to accept information. The ecology of information studies the relationship between informer and user of information, and that of certain categories of informers to one another and to information carriers. (LSCA)

10/65-22 **Automation des Geistigen** (The Automation of Intellect) Hans M. Haldenwanger, NADO, *15* (2) June 1964, pp. 65-73. References.

The application of computers to work in every type of library requires a new generation of information officers in the industrial nations, who will learn to live with the machine. The relationship between the intelligence of the machine and the creative thinking of man is examined. The computer must be used by the creative worker to eliminate fruitless avenues in his research. Full use of the computer in logical and matehematical calculation releases the scientist from time-consuming checking of results. (LSCA)

10/65-23 **Nutzen wir die Informations-Explosion richtig?** (Do We Utilize the Information Explosion Correctly?), Gary H. Hiller *Chemiker-Ztg. 88,* (5), 155-159 (March 5, 1964). No. (C.L.B.)

10/65-24 **Librarianship and Information Science,** G. Jahoda, *Florida Libs., 14* (4) December 1963, 7-10.

Librarianship in science and technology has not kept pace with the growth of scientific and technical literature or with the demands made on it in this field. Information science is slowly developing to meet the need. Some special libraries in industry come close to qualifying for the description of information centres, being those that prepare abstracts and translations, perform literature searches, and carry out annotating and indexing work. Basically there is a gulf between the librarian and the information scientist, and to bridge this plans are being made in Florida for the training of students in information science which, in addition to the basic librarianship training, will include techniques of handling information which will emphasize nonbook material. Related courses on psychology, linguistics and business administration will also be available. Students recruited will have good science backgrounds. As information science develops, librarianship will contribute to its fundamental principles and theory, while the benefit of a fresh view-point will be available to the librarians. (LSCA)

10/65-25 **Erfahrungen und Erkenntnisse bei der Aufstellung von Thesauren** (Problems and Solutions in the Making of Thesauri), Karl-Christoph and Malve Rothkirch-Trach, NAD, *15,* (3) September 1964, 118-121. References.

The construction of thesauri is considered from the viewpoint of pure logic. Words in the library have two functions: (a) enable information retrieval to be achieved by computer storage to express the query and answer to a problem verbally, (b) to methods. The computer deals with logical mathematical symbols and the words in the thesaurus serving it must not be ambiguous. A machine cannot recognize verbal shades of differences in a word, because each word-unit must have a clear cut logical and constant meaning. This leads to a special problem in German, a language geared to word building. It is necessary to break down compound words and use single terms or concepts as descriptions. Mechanical storage equipment can only operate by being fed narrowly defined single word units. (LSCA)

10/65-26 **Two Notes on the Quality of Economic Data,** Kryka M., *Mechanizace a Automatizace Administrativy* (Mechanization and Automation in Administration), No. 7, 1964, pp. 174-176. In Czech.

Quality requirements for economic data that are often neglected in practical work are discussed. It is necessary that such data be reliable, exhaustive, and available when needed. Criteria for determining the time factors and the reliability of such data are proposed. It is emphasized that any deviation of information data from the accurate values must be within the pre-established tolerance limits. The author lists consecutive stages in administrative procedures involving equipment, manpower, training and testing operations which, together with regular reports from individual sectors, can ensure smooth information processing. Cooperation based on an accurate definition of duties and responsibilities should result in high-quality economic data. (M.H.G.)

10/65-27 **Beware the Tax Collector.** Leslie P. McCarty, *SCIE 147*, 986 (1965).

A letter discussing the deductibility for tax purposes of subscriptions to technical journals by scientists. (K.G.)

10/65-28 **The Concept Library: Containing the Information Explosion,** Robert B. Miller, *1964 IEEE International Convention Record,* Part 10, 74-82 (August, 1964).

The increasing pressures for interdisciplinary knowledge in innovation and in technical and managerial decisions cannot be met by traditional practices in education and library service. Frequently it is sufficient to know the central working concepts and principles in a new field of science, technology or management for the development of hypotheses, the making (or withholding) of project evaluations, or for instigating literature searches in unfamiliar domains. Using insight, many technical topics can be abstracted into less than a dozen major principles, variables and working ideas. With further care and insight, a defining concept or principle can be communicated to the educated layman by test and diagram in a single page. Each page has bibliographic and index terms for more detailed reference. Many individual pages maye be recombined into different sequences for different needs and contexts. A library of thousands of such pages covering many technologies can be envisioned. Aside from snap briefings in new fields, or as quick review of once-learned knowledge, such a file of concepts could be made to serve as a dynamic link between the information dissemination and retrieval functions of an organizational library and the teaching functions of the educational facilities.

10/65-29 **The Growth of Scientific-Technical Literature,** V. Nemec, *Technicka Knihovna* (Technical Library), No. 9, 1964, pp. 267-273. In Czech.

Using the data from *Chemical Abstracts, Chemisches Zentralblatt,* and other sources, the author presents figures on the growth of scientific-technical literature in the world. Articles published on biochemistry, chemistry, and biology are tabulated by major periodicals, language, year of publication, and country of publication. The changes in the research and publication activities of various countries are analyzed, and the tremendous increase in the volume of scientific literature is shown. (DOC. INC.)

10/65-30 **Needless Pains Caused by Heedless Editors,** Irvine H. Page, *SCIE 147,* 1241 (1965). A letter deploring two practices of editors of scientific publications: the passing on to authors of "sanctimonious, snide, and picayunish alterations and criticisms made by referees," and frequent changes in the format of bibliographies (K.G.)

10/65-31 **The Machine and the Librarian,** Winter 1965, Ralph H. Parker, LRTS 9(1): 100-103.

10/65-32 **Operacio-kutatasi modszerek alkalmazasa a muzaki tajekoztatas teruleten** (The Employment of Operation-Research Methods in Technical Documentation), Ferenc Patek. *OMK,* (2) 1964, 1-89. Diagrs., references. (Summaries in English and German):

Investigations based on statistical surveys and the plotting of curves show that the average requirement level for scientific information is supplied by the interaction of motivation and inhibition factors. The obstructing factors are: shortage of time, lack of professional literature (to a small extent, 4%), lack of knowledge of foreign languages. An optimum curve for the use of documents shows that information is obtained in a proportion of between 5 and 10%. Classifying, abstracting and translation services were much used. At the lowest level there is a tendency to use documents rather than primary professional literature, and at the highest level, to appreciate translation. Use of documents depends on the economic aspect. Some 10-12 periodicals contained 50% of the information required. Inquiries are most often at a general level; only a few people make use of literature searches. And 90% of time spent actually reading or processing material. Hungarian professional periodicals are not as good as those in other countries. (LSCA)

10/60-33 **Library Technology Project—today and tomorrow,** Gladys T. Piez. *Libri, 14* (4) 1964, 330-336. Bibliog.

Since 1959, the American Library Association has been seeking ways of implementing modern technology and scientific management in solving some of the administrative problems of librarians. This project, known as L.T.P., is financed through the Council on Library Resources, Inc., of Washington, by the Ford Foundation. Equipment and services already tested include documentary reproduction, floor coverings, furniture, circulation control methods, library insurance, adhesives, pamphlet boxes and shipping containers for books. A three-year programme has been organized by the ALA and Special Libraries Association to conduct research into library binding. L.T.P. works through the American Standards Association to establish standards for library supplies, and an example of its success is the agreement of four government agencies to standardize the size of mocrofiche technical reports. A technical information service is maintained for librarians throughout the world, answering individual enquiries and publishing informational reports. A loose-leaf publication called *Library Technology Reports* is to be published bi-monthly. The ALA has arranged for the establishment in March 1965 of the Office for Research, with L.T.P. as the nucleus, and it is hoped that L.T.P. will become financially self-supporting. Future plans are concerned with book-processing, book-storage, library buildings and staff-training. Valuable contacts are made with European librarians to keep abreast of current trends. (LSCA)

10/65-34 **Information Science-Fiction or Fact?** Frederik Pohl, AMDO, vol. 16, No. 2 April, 1965, pp. 101-104.

10/65-35 **Research on the Historical Origins of Scientific Documentation,** I. Polzovics, *Orszagos Muszaki Konyvtar es Dokumentacios Kozpont—Evkonyv 1962* (Hungarian Central Technical Library and Documentation Center, 1962 Yearbook), Budapest, 1963, pp. 232-298. In Hungarian.

There is not a single monograph of importance on the history of scientific documentation. The prevailing opinion is that scientific documentation began with the publication of *Pharmaceutisches Centralblatt* (Central Pharmaceutical Gazette) in Leipzig, Germany, in 1830. The author investigated this problem and discovered that *Journal des Scavans,* published in Paris in 1665, should be considered the first abstract journal in the world. There was a need to inform the industrialists and economists of France about the developments in other countries. The initiator of this publication was Jean-Baptiste Colbert. The year 1830 marks only the beginning of a differentiation and specialization in scientific documentation. The author also describes all the other abstracting services and information journals which appeared in the period between 1665 and 1830; analyzes the social, economic, and historical factors which contributed to their appearance; and investigates the impact which these publications had on the development of science and technology. He recommends that January 5, 1965, be internationally commemorated as the date on which scientific documentation and information was started with the publication of the first abstract journal. (DOC. INC.)

10/65-36 **Information Protection: Why? How?** On-line computers and data communications handle information that needs to be protected. This confidential information is stored in a computer's memory files which many people have access to. Confidential messages go out over easily tapped communication lines rather than in sealed envelopes. Thus unauthorized access to private information is a growing danger. Describes three ways confidential information can be stolen and tells how to prevent it. Eugene E. Sarafin. CENG, v. 12, May 1965, pp. 105-107. (BATR)

10/65-37 **$F^1a^1c^1e^2$ $I^2t^3s^1$ $N^3o^1r^2m^1$,** Ralph R. Shaw, AMDO, vol. 16, No. 2, April, 1965, pp. 77-80.

The title of this paper, as indicated above, is *Information Centers*. While it appears to read *Face Its Norm* with numerical superscripts on all of the letters, it should not be too difficult to write a computer program which would recognize the superscripts as indications of the frequency with which each letter is to be written, and then, by using all permutations and combinations of these letters, in the frequency indicated, to come up with *Information Centers*, in a relatively short time, as one of the possible choices. The program could, of course, be refined so that we would not actually have to write the superscript numeral 1, since we could instruct the machine to consider a blank space between two letters as meaning the numeral 1.

The characteristics of information centers, as they appear from reviewing the published and unpublished literature and from personal observation, are the following:

1. They deal in science information.
2. The current concept limits them to narrow areas of specialization.
3. Each is closely associated with a group of working scientists in its area of specialization.
4. While they collect and organize and index artifacts that contain data, the common thread in all attempts to define their functions and to differentiate them from other agencies that perform some or all of the same functions, is that their prime function is the application of a high order of judgment by scientists who have the highest possible substantive competence in the particular area of specialization to organize, summarize, abstract, evaluate or otherwise manipulate this raw material, retrieving from it the essence of information pertinent to the particular problem at hand that is contained in a large corpus of raw informational materials, and to communicate this essential knowledge to their peers.
5. They are designed and operated by scientists for scientists.
6. They use (or allegedly use, or might sometimes use) some form of hardware.

While the characteristics enumerated are of varying degrees of importance, and the stated objective may vary widely from the norm of actual operations or achievement, there can really be no difference of opinion about the need for improving information services to science, technology, and to all other areas of scholarship as well. Furthermore, whether we agree in whole, in part, or not at all, on any particular device selected for achievement of this goal, we must respect and admire the imaginative and courageous efforts that are being made to do something about this important social need.

10/65-38 **Osnovi Poznavan'a Kn'ige** (Fundamentals of the Science of Books), J. Shishkovitch, Publ. of Textbooks of the Socialist Republic of Serbia, Belgrad, 1963, p. 159; Reviewed by Shulagitch, R., in *Bibliotekar* (The Librarian), No. 3/4, 1964, pp. 221-223. In Serbo-Croatian.

The publication reviewed is a textbook for library schools in Yugoslavia. It discusses the development of writing and printing, the first written records in the world, manuscripts, the basic elements of modern books, printing methods, the preservation of books and their protection, bibliographies, and related fields. A dictionary of library science terms is included. The textbook is recommended for all students of library science in Yugoslavia. (DOC. INC.)

10/65-39 **Objectives and Standards for Special Libraries,** Professional Standards Committee, SPLB, 55 (10), 672-680 (Dec. 1964).

The product of a number of years study by the committee, this report offers a guide to the elements which must be present in a successful special library or information center. Outlined are the objectives of such a function and its requirements as to staff, acquisition policies, services offered and physical facilities. (M.C.P.)

10/65-40 **Dokumentation und Forscher als ihre Benutzer** (Documentation and Its Use for Research) Elin Tornudd. NADO, 15 (4) December 1964, 182-186. References.

Gives a survey of studies regarding information needs and uses made in the USA, the UK, and in Scandinavia and suggests that similar studies should be made in Germany and other countries. Extensive studies of documentation needs are necessary for the formulation of national and international policies; local studies will help in the development of a single documentation service. The methods used in these use studies are discussed and some results are compared. The functions of documentation in facilitating the scientists' approach to information are dealt with. (LSCA)

10/65-41 **Report on the Activities of FID in 1964,** W. Van der Brugghen, W., REDO, vol. 32, no. 2, May 1965, pp. 59-65.

10/65-42 **The Communication of Information,** Herbert Coblans, *THE SCIENCE OF SCIENCE,* Maurice Goldsmith and Alan Mackay, eds., London: Souvenir Press, 1964, 93-101. References.

At the Royal Society Information Conference 1948, J. D. Bernal, unsuccessfully proposed a scheme to abolish all existing scientific periodicals and replace them with central agencies to distribute single papers linked with abstracts on cards. In 1960 he again proposed the abolition of scientific papers as a form of scientific communication saying that the secondary sources, e.g. abstracts, reports, tables, were more important. The Weinberg Report (1963) arrives at similar ideas, i.e. specialized information or data centres to supply information, backed up by large libraries in which the original documents will be found. To abstract and synthesize the information, scientists and engineers are needed because machines cannot do all the work. (LSCA)

10/65-43 **Libraries of the Future,** Licklider, J. C. R., Cambridge, Mass.: The M.I.T. Press, 1965, 219 p. $6.00 (L.C. 65-13831).

A more accurate if less profitable title for this work would have been "The Automated Information Center of the Future." As the author sees it, technological advances between now and the year 2000 ("the future") will make it possible for us to create a national network of computers containing all "solid" literature (all knowledge). Instead of using books or publications, the scientist, engineer, and scholar will use a console tied into this system by means of on-line, time-sharing techniques. The user will be able to carry on a rather sophisticated conversation with the computer using devices similar to input-output typewriters, oscilloscopes, light pens, advanced printout devices, and so on, to obtain all relevant information. Perhaps the computer can even respond to verbal requests. The "pages" and facts contained in the system will have been so well organized, precorrelated, and analyzed by the computer that segmentation or specialization of research will no longer be a problem. This dynamic system will have no need for books or physical documents, since refined search techniques, display screens, and rapid printout methods will make these passive artifacts superfluous. The creation of such a system is predicated on the economic assumption "that interaction with information and knowledge will constitute 10 or 20 per cent of the total effort of the society." The network will be justified on the basis of society's increased productivity and effectiveness resulting therefrom.

The second part of the book is a brief review of recent linguistic and computer research studies having possible application to such an advanced information network. As a bibliographic essay for communications and linguistic researchers unfamiliar with recent work on syntactical and semantical analysis, it has considerable value. A bibliography of some 40 papers in these areas is included.

The basic argument presented for this "network of the future" is hardly compelling, since only an abstract hypothetical system is presented with insufficient consideration of user needs, goals, and communication patterns. Not enough attention is given to the problems of index structure and machine indexing patterns. No consideration is given to advantages, disadvantages, and future use of microfilm, which apparently was dismissed out of hand. I would like to have seen a comparison of the proposed network organization to a system more closely resembling current practice but expanded through a corresponding expenditure of resources.

As you might suspect from the above, the major portion of this work has a certain resemblance to science fiction. It is stimulating to read, but the use of jargon and coined words, combined with a plodding style and loose organization, make it hard to follow. Do not expect to find here any immediate help to your present problems. Neither will you find assistance for planning existing or new special libraries. My guess is that at best this book will serve to create more interest in current linguistic research.

This work is based on a study conducted by Bolt Beranek and Newman, Inc. under contract to the Council on Library Resources Inc., between November 1961 and November 1963. The publishing of the research results in technical journals, listed in the bibliography, should prove to be of greater value than the book itself. (SPLB)

Published Reference Tools

10/65-44 **Thesaurus of Descriptive Terms and Code Book—** Bureau of Ships, Second Edition (Supersedes Navships 250-210-1; 1st Edition, Dec. 1963). Compiled by Bureau of Ships Technical Library. NAVSHIPS 0900-002-0000 (Formerly: Navships 250-210-1), March 1965, Bureau of Ships, Navy Department, Washington, D. C.

The second edition of the Bureau of Ships *Thesaurus of Descriptive Terms and Code Book,* developed for use in conjunction with Project SHARP, contains approximately 4,600 main terms, an increase of nearly 100 new terms. Effort was concentrated on revising existing terms to display more realistic hierarchal and general relationships, inclusion of additional scope notes, and deletion of some outmoded and obsolete terms.

In 1964, plans were formulated to computerize the Thesaurus utilizing the UNIVAC-LARC computer at the Applied Mathematics Laboratory of the David Taylor Model Basin. It was felt that mechanization of the Thesaurus would facilitate:

(1) automatic posting of generic terms,
(2) automatic generic searching,
(3) automatic publication of the Thesaurus, and
(4) general updating.

The Thesaurus is divided into three sections: the Thesaurus proper with all terms and cross references; a code book arranged alphabetically by term; and a code book arranged alphabetically by code word. Code words are not included in the main body of this edition.

10/65-45 **Thesaurus of Descriptive Terms and Code Book,** Camp, Ruth D., Bureau of Ships Technical Library, Department of the Navy, Bureau of Ships, Washington, D. C. 20360, December 1963, first edition, (NAVSHIPS 250-210-1).

10/65-46 **An Annotated Bibliography of Publications on Dependency Theory,** Rand Corp., Santa Monica, Calif., by David G. Hays. March 1965, 22 p. Rept. no. RM-4479-PR. AD-613 469, Div. 32. CFSTI Prices: HC $1.00; MF $0.50. Contract AF49 638 700. Unclassified report.

The books and papers listed in the bibliography have in common a concern with the notion that syntax is best described by specifying word-to-word connections, generally called "dependencies," rather than by segmentations of sentences. Some of the works listed contribute to the development of a formal dependency theory in linguistics. Others apply the growing theory to the description of natural languages and to the design of computer systems for machine translation, information retrieval, and other purposes. A few of the papers cited criticize and reject the dependency notion.

10/65-47 **Computational Linguistics: Bibliography, 1964.** Rand Corp., Santa Monica, Calif., by David G. Hays and Roxana Ma. March 1965, 60 p. RM-4523-PR. Contract AF49 638 700. AD-613 311, Div. 32, 30. CFSTI Prices: HC $3.00; MF $0.75. Unclassified report.

Contents (809 items): Classification theory, computation and programming, computational linguistics, computers and hardware, documentation, linguistics, non-numerical computer applications, psycholinguistics, author index and organization index.

10/65-48 **Air Force Office of Scientific Research Information Sciences 1964.** Harold Wooster, Director, Directorate of Information Science Office of Aerospace Research, United States Air Force. AFOSR 65-0271, March 1965.

"Information Sciences," as a generic name for the body of scientific disciplines and understanding underlying the technology of information handling, is perhaps no more than five years old. Like any member of the *nouveau riche* we have been searching our family tree for named, and unhung, ancestors.

One of these ancestors is the abstract journal—just 300 years old this year. On January 5, 1665, "Les Journal des Scavans" was established. This was a weekly publication created to abstract scientific and literary books of contemporary Europe, reviewing new inventions and discoveries in physics, chemistry, astronomy, and anatomy. This abstract journal preceded by two months the first journal of primary publication, the "Philosophical Transactions" of the (British) Royal Society.

Not inappropriately for the information sciences either, the second editor of this journal was the cleric, Jean Galois, an expert in physics, mathematics, astronomy and linguistics. The study of heavenly bodies is the only of Galois' fields not of interest to the modern information scientist; item 2.4 in the text concerns the application of one of Galois' discoveries to modern documentation.

Another ancestor of today's information sciences is the digital computer. This can be traced at least as far back as the Leibnitz plea used as epigraph. One can certainly trace a more or less direct line back to another clergyman, George Boole, and his publication in 1847 of *The Mathematical Basis of Logic*—even though this had to wait for 90 years until Claude Shannon's M.S. thesis of 1937, on the application of Boolean algebra to switching circuits.

Babbage, long before the days of government R & D contracts, died almost in poverty in 1871. He had devoted the thirty previous years to building his Analytical Engine, devoting almost superhuman ingenuity in trying to do with gears and cams, building the machines to build his machines, mechanically things that had to wait for the discovery of the Edison effect, the de Forest tube, and eventually, the transistor. Three years after Aiken started building Mark I in 1939 he discovered Babbage, and found that he was going well and easily down the same paths Babbage had traversed with difficulty.

Enough of ancestors. What of the information sciences today? There is a rapidly developing *technology* of information handling—a technology concerned with the handling of ever increasing amount of scientific and technical, logistic, intelligence and command and control information. It is axiomatic that advances in any technology can only come about in three ways:

by specific research and development in that technology *per se,*
by exploiting the fortuitous advances in other, related technologies and
by advances in fundamental scientific knowledge and understanding.

This is a report on the research program sponsored by the Directorate of Information Sciences of the Air Force Office of Scientific Research during 1964. This Directorate is just one of the Scientific Directorates of AFOSR. Like the others, it is primarily concerned with selecting and supporting long range basic research—in our case, in the information sciences. Unlike the other Directorates, it is also involved with a particular aspect of technology—the technology of information processing, as it applies to current and predictable AFOSR operating problems.

The research program reported herein is a resultant of these two forces. About one-third of our program is devoted to research in the technology of information handling; two-thirds to the information sciences. The text of this report has been written by our investigators; editorial changes in most cases have been confined to elision. We hope that the resulting diversity of styles will help to show the diversity and challenge of their intellectual quests.

It would be impossible to administer this research program without the help of my two associates, Rowena Swanson and Thomas K. Burgess, Capt. USAF, nor without the support given by the Executive Director of AFOSR, Dr. William Price.

10/65-49 **Computer Literature Bibliography 1946-1963,** by

W. W. Youden, National Bureau of Standards Miscellaneous Publication 266; March 31, 1965; 463 pages; $3.75. (Order from the Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. 20402, or from local U. S. Department of Commerce Field Offices.)

Comprehensive, rapid, and economical access to periodical and proceedings literature in specialized subject areas has been an increasing problem in recent years. This computer-produced bibliography and index is intended to provide such access to the more than 6,000 articles published in 9 journals and over 100 proceedings in the field of computers and data processing. This bibliography is completely comprehensive for the selected journals and proceedings. Every article is listed with its full title and all authors.

The Title Word (KWIC) Index section of the bibliography enables the reader to find an article if any part of its title is known, or to find all the articles whose titles include a particular word or phrase. The Author Index section lists all authors of each article along with the titles of their articles. Both the Title Word and Author Indexes will guide the user directly to the desired article without a cross reference or second look-up.

This bibliography will be a valuable tool for all computing and data processing specialists, and a necessity for technical libraries that cover any part of the computer field.

10/65–50 **Monograph Cataloging Manual.** AD–613 337, Div. 32. CFSTI prices: HC $4.00; MF $1.00. Air Force Cambridge Research Labs., Bedford, Mass. June '64, 117 pp. Unclassified report. See also AD–435 615.

Temporary instructions, issued as each phase of the overall plan is accomplished and tested, will eventually be published for Serials Control and Cataloging, Acquisitions (Order and Receipt) Binding, and Special Routines (including Computer Programs). The Monograph Cataloging Manual serves to: (1) Set forth for the mongraph cataloger the authorities and practices adopted by the Library in cataloging its technical publications. (2) Set forth for the machine room operators the machine configurations, encoding formats, and output requirements of the automated system. (3) Interrelate the functions of both cataloger and data processor in description of work flow from point of cataloging through subsequent processing operations (generation and filing computer tapes, book preparation, and preparation of accessory records) to the point fo circulation and reference activities.

10/65–51 **List of Recent Reports of Interest to Technical Information Retrievers.** Compiled by Richard H. Austin and John Cook Wyllie. 35 pp.

Compiled as part of a continuing literature search for the Industrial Research and Development Center of the School of Engineering and Applied Science, this selective cumulation is issued in April 1965. Permission to duplicate it for any reason is granted. Science Reference Division. University of Virginia Library, Charlottesville, Virginia 22903. Phone: 703-295-2166. TWX: 510-587-5453.

10/65–52 **Consolidated List of Information Retrieval Words.** AD–436 827, Div. 32, 31. OTS price $14.50. Bureau of Naval Weapons, Navy Dept., Washington, D. C. 1960, 214 pp. NAVWEPS LS 3. Unclassified report.

10/65–53 **Root-Term Glossary for the Subject Analysis of Technical Reports, with Computer Codes for Use with the NOL 7090 Retrieval Program.** AD–436 843, Div 32, 22. OTS price $10.50. Bureau of Naval Weapons, Navy Dept., Washington, D. C. 1960, 138 pp. NAVWEPS LS5. Unclassified report.

10/65–54 **Root-Term Vocabulary for the Subject Analysis of Technical Reports in Aeronautical and Related Fields in the Technical Library Collection.** AD–436 844, Div. 32, 1. OTS price $6.60. Bureau of Naval Weapons, Navy Dept., Washington, D. C. Aug. 60, 61 pp. NAVWEPS LS 4. Unclassified report. In cooperation with Engleman and Co., Inc., Washington, D. C., Contract Noas59 6243c.

10/65–55 **COSATI Subject Category List.** AD-612 200, Div. 32. CFSTI prices: HC $1.00; MF $0.75. Federal Council for Science and Tech., Washington, D. C. Dec 64, 60 pp. PB 166 877. Unclassified report.

The Subject Category List presented was endorsed by the Committee on Scientific and Technical Information (COSATI) of the Federal Council on Science and Technology as a uniform subject arrangement for (1) the announcement and distribution of scientific and technical reports which are issued or sponsored by Executive Branch Agencies, and (2) for management reporting. The list is a two-level arrangement consisting of 22 major subject fields with a further subdivision of the fields into 178 groups. Scope notes are included for each group. Abstracts, citations and the like, for announcement purpose, can be gathered into these broad subject fields or groups for display to the user. For distribution purposes, these fields or groups may likewise be employed. Similarly, the fields or groups may be useful for arranging projects, tasks, or programs for management reporting purposes. (author)

10/65–56 **Resume of U. S. Patents Obtained Fiscal Years 1961-1963 by Office of Patent Counsel U. S. Naval Missile Center.** AD–436 585, Div. 32, 4, 7. Naval Missile Center, Point Mugu, Calif. 10 Apr. 64, 32 pp. NMC MP 64 3. Unclassified report.

This report of U. S. Patents has been prepared by the Office of Patent Counsel for the purpose of making readily available this unclassified technology to all government personnel. The single drawing and descriptive matter from each patent has been copied directly from the weekly publication Official Gazette of the United States Patent Office, U. S. Department of Commerce. The U. S. Government either has title to these inventions, or has a royalty-free license to use these inventions for governmental purposes only. (author)

10/65–57 **IDEP to DDC Report Number Cross-Reference Index.** AD–435 650, Div. 32. Sylvania Electric Products, Inc. Waltham, Mass. 25 Mar. 65., 38 pp. Unclassified report.

This index is for use in determining the Defense Documentation Center Report Number when the IDEP Report Number is known. All IDEP reports announced in DDC's Technical Abstract Bulletins (TAB) up to and including February 1, 1964,, are listed. (author.)

10/65–58 **Science-Technology Subject-Heading Lists in English, Recent or Still Useful in Programming Information Retrieval Systems.** John Cook Wyllie, Librarian of the University of Virginia.

This list was first issued in October 1964 as a literature search for the Industrial Research and Development Center of the School of Engineering and Applied Sciences. This second issue, with four more titles, is issued in April 1965. Permission to duplicate it for any purpose is granted. 8 pages. Science Reference Division, University of Virginia Library, Charlottesville, Virginia. 22903.

10/65–59 **A New Look At Feature Card Indexing: A Symposium on Aspects of Practical Application, 8th April 1964.** London, Dr. W. E. Batten, Chairman. ASLP 16 (11) November 1964, 333-340.

Ten authors describe the systems operating in their own organizations and comment on their effectiveness. In total the papers provide an assessment of feature card indexing and its various uses. (LSCA)

10/65–60 **Die Dokumentation der Wasserwirtschaft,** Documentation of Water Undertakings), Helmet Hubner. NADO, 15 (4) December 1964, pp. 187-191.

There is an ever-increasing scientific interest being taken in the use of water to meet the various requirements of household, trade, industry and agriculture. The German Working Committee relating to Water Researches has undertaken the task of compiling and publishing *Documentation Water*. Arrangements are made with agencies and scientific institutions in Germany and abroad to compile summaries of noteworthy papers on water and by this method a complete bibliography can be compiled. *Documentation Water* has been published monthly since October 1960 and contains meanwhile 36 leaves with 288 items. (LSCA)

10/65–61 **Library Statistics 1962.** Published by Zentralinstitut fuer Bibliothekswesen, Berlin, reviewed in *Zentralblatt fuer*

Bibliothekswesen (Central Newspaper of Library Science), Vol. 5, 1964, pp. 283-284. In German.

This is a collection of statistical data on all library activities in East Germany: number of readers, number of books checked out, new accessions, etc. The amount of the data reported increased considerably in comparison to the 1961 volume. It is shown that the number of regular readers and the number of books checked out are growing steadily. In the report year, 18.7% of the population of East Germany were registered users of libraries. (DOC. INC.)

10/65–62 **Abbreviaturae Cyrillicae,** (Library Abbreviations in Cyrillic Script, E. Moravek, and I. Weger, Budapest, MTA Konyvtar, 1961, 138 pp. Vol. 1. Vocabularium Abbreviaturarum Bibliothecarii. Reviewed by H. Hohne, in *Zentralblatt fuer Bibliothekswesen* (Central Newspaper of Library Science), No. 1, 1964, pp. 45-47. In German.

This is the first volume of a vocabulary of abbreviations in Cyrillic characters used in library and documentation literature. Russian, White Russian, Bulgarian, Macedonian, Serbian, and Ukrainian abbreviations are listed alphabetically. Full texts of the abbreviations are followed by a translation into Hungarian, German, and English. The reviewer points out a number of inaccuracies in the German and English translations of abbreviated terms. Volume II of the vocabulary will be a *List of Bibliographical Abbreviations in Latin Characters*; Volume III, a *List of Abbreviations of Names of Institutions and Companies in European Languages.* (M.H.G.)

10/65–63 **Reprography and Copyright Law.**—Proceedings of a symposium in 1963 sponsored by The American University. Lowell H. Hattery and George P. Bush, editors, 204 pp. 1964. American Institute of Biological Sciences, Washington, D. C. Explores the reprography-copyright problem, its varied interests, viewpoints, proposed solutions, and outlook for the future. (BATR)

10/65–64 **Searching the chemical literature.** M. G. Mellon, Washington, D. C., American Chemical Society, 1964. 40 pp.

This is an introductory guide to the literature of chemistry. The topic is dealt with in three sections, namely: 1. What are the questions that a chemist asks a library? 2. Where are the answers? 3. How are the answers found? In section two the sources are divided into primary, secondary, and tertiary. Primary sources comprise the publications which are devoted almost entirely to new material, e.g. periodicals, government publications, patents, theses, house organs, and other trade literature, separate volumes of papers, preprints of papers. Secondary sources comprise the publications which serve to collect, organize, and disseminate the information in primary sources, e.g. text books, tabular compilations, dictionaries and encyclopedias, monographs, treatises, formula indexes, review serials, abstracting journals, indexing serials, bibliographies. Tertiary sources comprise guides to primary and secondary sources and publications providing facts about chemicals, their organizations and their products, e.g. guides to the literature, periodicals on documentation, biographical works, scientific aid, technical society directories, book lists, language dictionaries, books on laboratory planning and administration, directories and trade catalogues and periodicals in financial data. In section three the publications to consult in current awareness and in retrospective searches is discussed. (JDOC.)

10/65–65 **Metodyka Bibliograficzna,** (Methods of Bibliographical Work. Edited by H. Hleb-Koszanka, M. Dembowska, and H. Sawoniak. Second edition, Polish Library Association Press, Warsaw, 1963, pp. 402. Reviewed by J. Kossonoga in *Przeglad Biblioteczny* (Library Review), No. 1/2, pp. 89-92. In Polish.

This new edition of the handbook describes various methods and tools for the preparation of bibliographies. It covers such topics as title entries, indexes to bibliographies, typographic considerations in printing bibliographies, special bibliographies, local and regional bibliographies, etc. Most chapters have been reviewed and updated, thus incorporating the latest ideas and developments in this field. (DOC. INC.)

10/65–66 **Data Processing Yearbook.**—As with the previous Yearbook, there are four major areas of emphasis: Tools, Techniques, Tactics, and Reference Information. Includes the directory of data processing at colleges and universities, which shows an expansion to nearly a thousand digital and analog computers in use on the North American continent. Directories of manufacturers, service organizations, and audio-visual aids are included.—Edith Harwith Goodman, editor. 355 pp. 1964. American Data Processing, Detroit, Michigan. (BATR)

10/65–67 **Organic Name Reactions.** Helmut Krauch and Werner Kunz. John Wiley (New York), 1964. XXIV+620 pages, Illus. Reviewed in *SCIE 147,* 726-7 (1965). (*K.G.*) $16.00.

10/65–68 **The DDC Thesaurus: Past, Present and Future.** Paul H. Klingbiel. ASLP, 16 (8) August 1964, 252-257. References.

A brief historical introduction is given and the present structure of the thesaurus is outlined. It is indicated that a UNIVAC 1107 thin-film computer will be an important tool in the future vocabulary building process. Due to changes in U. S. Government procedure and co-operation with the British Ministry of Aviation substantial modifications are now taking place in the vocabulary. The author mentions the possibility of merging a new edition of the thesaurus into a standard National Thesaurus. The considerable problems facing any National Thesaurus are then discussed. Retrieval at DDC has shown that in making a search users do not use a large portion of the total indexing vocabulary. As computer time it may be economical to withdraw low-frequency use terms for manual search. (LSCA)

10/65–69 **Encyclopedia of Polymer Science and Technology. Vol. 1; Ablative Polymers to Amino Acids.** Herman F. Mark, Norman G. Gaylord, and Norbert M. Bikales (Eds.) Interscience, Wiley, (New York), 1964. XVIII+893 pp. Illus. $50 (subscription price, $40 per volume). Reviewed in *SCIE 147*, 386-7 (1965. (*K.G.*)

10/65–70 **The New Mathematics Dictionary and Handbook.** Robert W. Marks. Bantam Books (New York), 1964. 186 pages. 95 cents. Reviewed in *SCIE 146,* 1670 (1964). (*K.G.*)

10/65–71 **Standardization in the Field of Bibliography and Reproduction Techniques.** J. Pelcowa, *Biuletyn Instytutu Bibliograficznego* (Bulletin of the Bibliographical Institute), No. 3, 1963, pp. 189-253. In Polish.

The article gives general information about standardization, the terminology used in this field, and the history and organization of standardization of bibliography in Poland. Established standards of bibliography are listed and also a short description of the five year plan for 1961-1965 in the field of bibliography and reproduction techniques. Standardization problems of this particular branch such as standards concerning editorial composition of publications, reproduction of documents, etc., are discussed. International exchange and cooperation is handled through the International Organization for Standardization (ISO), whose actual work is done by technical committees. Poland, besides conducting research on the national scale, participates in international programs, mainly those conducted by the Technical Committee No. 46 of the ISO. Finally the article lists selected bibliographies in the field of standardization. (T.T.)

10/65–72 **Analysis of Library Systems: A Bibliography.** L. Pratt. SPLB 55 (10), 688-695 (Dec. 1964). An extensive list of references on systems design, time and cost studies and evaluation techniques. (M.C.P.)

10/65–73 **Publications of the Documentation Institute of the Polish Academy of Sciences.** G. Rozsa, *Tudomanyos es Muszaki Tajekoztatas* (Scientific and Technical Information), No. 3, 1964, pp. 185-188. In Hungarian.

In a short informative article the author reviews the most important publications (serials and monographs) of the Documentation Institute of the Polish Academy of Sciences since its establishment. Of special interest to the English-reading public might be the publication *Polish Scientific Periodicals,* which contains abstracts of articles and includes an author index. (DOC. INC.)

10/65–74 **An Annotated Bibliography of Studies on the Flow**

of Medical Information to Practitioners.—The 120 annotated references are indexed according to; the communication channels evaluated, what was measured, and which techniques were used.—Andrew M. Sherrington, compiler. *Methods of Information in Medicine,* v. 4, no. 1, Mar. 1965, pp. 45-57. (BATR)

10/65–75 **Directory of Microfilm Readers and Microfilm Reader-Printers.**—Includes 29 manufacturers and characteristics of 104 readers for roll film, micro opaques, aperture cards, microfiche, and sheets.—SSTM, v. 6, no. 3, Apr. 1965, pp. 45-49. (BATR)

Characteristics of Information Resources

10/65–76 **Documentation and Dissemination of Research and Development Results.** Study No. IV. U. S. Congress, House of Representatives. Select Committee on Government Research. 87th Congress, 2nd Sess. Report of the Select Committees on Government Research . . . under the Authorty of H. Res. 504. Nov. 20, 1964. Washington, U. S. Govt. Print Off., 1964. 148 pp. USGPO price $0.60. (Committee Print)

Deals "with Federal scientific and technical information programs, their efforts to achieve efficacy and timeliness, their present scope, prevailing practices, access to and utilization of foreign information, problems that face them, and proposals for dealing with them."

10/65–77 **A World-Wide Future,** K. A. Baer. SPLB 56 (2), pp. 87-92 (Feb. 1965). A report on the development and present activities of the International Federation of Library Associations.

10/65–78 **Summary Papers and Summary Journals in Chemistry.** June 1965. J. D. Bernal. JDOC Vol. 21, No. 2, pp. 122-127.

The problems in keeping up with scientific literature by reprints is noted to have failed a decade ago. Perusal of research installment reports is equally wasteful of time. The verdict is not yet in on the system of a journal broken into any number of parts. Comments on these alternatives are put forth to help the Chemical Society as it inquires into methods of publication. Prof. Bernal's "qualitative opinions" are offered in his argument for multiple primary publication: one aimed at those in a narrow specialty and another of the same paper as those with ancillary interest, or one paper for a chemical synthesis and a second on the new technique used in the synthesis. The value of secondary and tertiary publications is confirmed. (D.C.W.)

10/65–79 **The Library in the Service of Scientific and Practical Agriculture,** Bocsever, A. M. (Director of the Soviet Central Agricultural Library), *Mezogazdasagi Konyvtarosok Tajekoztatoja* (Bulletin of Agricultural Librarians), Vol. 4, 1963, pp. 3-9. In Hungarian.

The Central Agricultural Library of the Soviet Union is located in Moscow with a branch office in Leningrad. Some 2.5 million books and periodicals are in stock, half of which are publications in 30 different non-Russian languages. The subject area encompasses all domestic and foreign literature pertaining to agriculture and biology. Part of the acquisition comes from international publications exchange with 50 nations and over 500 exchange organizations. The Central Agricultural Library directs also the activities of a nationwide system of over 1,000 local agricultural libraries. It holds over three million index cards and publishes a union subject catalog. One of the publications of the library is the *Agricultural Literature of the Soviet Union,* a bibliography averaging 60,000 domestic titles annually. A biweekly *Bulletin of Domestic Agricultural Literature* gives quick information on central and regional publications. Over 22,000 foreign articles are referenced in *Agriculture* (ten issues per year). Translations of foreign articles appear in *Agriculture Abroad.* The library also publishes a number of bibliographies. Plans are being made to mechanize and automatize the library, for which purpose computers shall find wide application. (B.C.L.)

10/65–80 **Some Prerequisites to Cooperative Cataloging,** Ritvars Bregzis. CRILI 25 (6) November 1964, pp. 497-500. References.

There is an urgent need for a co-operative cataloguing centre to process current non-American materials. Two alternatives have been suggested: centralized cataloguing, in which libraries would forward acquisitions to the centre for cataloguing; and co-ordinated cataloguing, whereby the centre would act as a clearing-house receiving and distributing requests for cataloguing to be performed by the library first acquiring a given item. Both plans have advantages and disadvantages, but, in either case, certain problems would have to be resolved before work could be undertaken. The first problem concerns the technicalities of conventional cataloguing; e.g. standardization of rules of entry, classification and subject heading. The second concerns the introduction of automation into catalguing; the basic systems of bibliographic control need to be re-examined to bring them into line with the logic and consistency of modern data processing. (LSCA)

10/65–81 **A Note on Cataloguing at the International Level,** Herbert Coblans. Reprinted from *Med boken som bakgrunn. Festskrift til Harold L. Tveteras,* Oslo 1964, pp. 197-201. Bibliog.

Problems of cataloguing at the international level have been brought to the fore with the publication of ICCP principles. These, however, are mainly concerned with choice of main entry, and could be adopted by all countries irrespective of language. An international library or one in a multi-lingual country has other problems which have not been systematically studied. The recent Manual published by the U.N. Library states that geographic names and corporate authors should be rendered in the language of the bibliography. On the other hand the FID considers English, French, German, Russian and Spanish as equal, and uses any of these without translation. The practical implications of this polyglot attitude contrasted with a favoured single language in the case of geographical location and corporate authorship are considered. The inclusion of Russian in FID's five languages also gives rise to transliteration problems. (LSCA)

10/65–82 **Zentralstelle fur maschinelle Dokumentation** (ZMD) (Central Office for mechanical documentation) Martin Cremer. NADO, 15 (4) December 1964, pp. 167-168.

In order to meet the growing demand for mechanical documentation in West Germany, the director of the Institute of Documentation Methods established in 1964 this Central Office of mechanical documentation (Z.M.D.). The work of ZMD is in three parts. (i) Advice is given to organizations wishing to install mechanical documentation systems. (ii) Study is made on programming methods. (iii) The installation of documentation machines were requested. (LSCA)

10/65–83 **Technical Libraries in the U.S.S.R.,** O. S. Cubar' jan UNDL 18 (5) September-October 1964, pp. 224-229, 242.

(i) The technical library network. In 1962 there were in the Soviet Union over 16,000 technical libraries (with a total stock of 175m. v.) serving industry, building and transport. Shows how state control of the economy has effected the planning of the network. (ii) Describes the work of the various types of technical library: the State Public Scientific and Technical Library of the USSR, which is at the head of the network: the "central" libraries, by which is meant the regional (or republic) libraries and those libraries which serve a particular industry on a nationwide scale; and libraries attached to scientific research institutes and planning and construction organizations. (iii) The main features of the work of technical libraries are listed and the author explains what uses are made of the printed-card services of the All-Union Book Chamber and other specialized organizations. Methods of information work; bibliographical bulletins, abstracting journals and groups, reviewing seminars, exhibitions and readers' conferences are described. A large number of volunteers assist at all levels in technical library work. (LSCA)

10/65–84 **Organisation des Dokumentationswesens auf nationaler und internationaler Ebene.** (Organisation of Documention Methods at National and International Levels) Ingetraut Dahlberg. NADO, 14 (2) June 1964, pp. 90-99. References.

National documentation and information centres of five European countries (France, Poland, Czechoslovakia, Hungary and the Federal Republic of Germany) are presented and their functions and missions towards subordinate documentation and

information services described. Organizational charts show mutual relation and interdependence at the different levels. A consideration of the systems described makes it appear advisable to give up centralization on a national level in favour of a centralized survey of specialized documentation centres covering merely one single subject field each. Such centres should only abstract the literature of their own country and exchange it with that of other countries. Such international exchange should be controlled by Unesco, the International Federation of Documentation (F.I.D.) and the International Association of Documentalists and Information Officers (A.I.D.). (LSCA)

10/65–85 **Documentation of Special Agriculture Literature in the German Democratic Republic,** Fabinyi, R., *Mezogazdasagi Konyvtarosok Tajekoztatoja* (Information on Agricultural Libraries), No. 1, 1964, pp. 9-12. In Hungarian.

Information on agricultural literature is processed by the Information and Documentation Institute of the German Agricultural Academy and special local information services of agricultural research institutions. The academy publishes the monthly reference journal *Landwirtschaftliches Zentralblatt* (Central Journal on Agriculture) which contains sections on agricultural techniques, plant production, animal husbandry, and veterinary medicine. The preparation of a monthly catalog that will be published by the academy is discussed. The documentation work of research institutions includes: the publication of a subject catalog on botany, genetics, and plant protection in Gross Luesewitz; the preparation of reports on new research endeavors that appear in periodicals; and the distribution of photocopies of studies in new research problems to all research institutes. (DOC INC)

10/65–86 **The Council of Rectors of the Chilean Universities and the National Center of Information and Documentation (CENID),** Cesar Fighetti S. *Libri,* 14 (4) 1964, 337-344.

The rectors of the eight universities of Chile are jointly responsible for the administration of a national fund for universities, which is spent in various ways, including special projects. One of these projects is the National Center of Information and Documentation, known as C.E.N.I.D. Although libraries exists in Chile are not generally well-developed, good collections exists in universities, government departments and private organisations, and the task of C.E.N.I.D. is to co-ordinate the work, as well as to encourage the development of these libraries and advise on the training staff. The C.E.N.I.D. staff is limited to a director, his assistant, clerical staff, and two operators of reproduction equipment. Originally limited to the sciences, the scope of the organisation is to be extended to cover the humanities. Additional finance and technical advice has come from the United States, and the advising scientist, Dr. Igelsrud, holds that the centre must maintain contact of a direct, informed and sympathetic nature with the individual scientist requiring information, if it is to obviate his having to conduct literature searches personally. The centre is compiling three guides for research: a union catalogue of scientific and technical periodicals; a guide to scientific and technical research in Chilean universities (using the British publication *Scentific research in British Universities and Colleges* as a reference), and a guide to Chilean libraries. Another American expert specialises in documentation and librarianship and assists with the practical administration of individual libraries, as well as advising on the future development of C.E.N.I.D. (LSCA)

10/65–87 **Work Organization in the Photolaboratory of a Documentation Center,** Gara, A., *Tudomanyos es Muszaki Tajekoztatas* (Scientific and Technical Information). No. 3, 1964, pp. 189-195. In Hungarian.

The variety of tasks performed in the photolaboratory of a documentation center requires a well-organized coördination of work. Usually these laboratories are affiliated with an enterprise or agency which determines their operating policies and procedures. The author discusses the problems of copying and processing documents with high efficiency as well as the problem of making maximum use of equipment. The work done outside the laboratory is analyzed, too. Procedures for production control and materials inventory are reviewed, and safety measures used in photolaboratories are investigated. (author)

10/65–88 **Technical Documentation in the German Democratic Republic,** Herpay, B., *Mezogazdasagi Konyvtarosok Tajekoztatoja* (Bulletin of Agricultural Librarians), Vol. 4, 1963, pp. 10-13. In Hungarian.

Development and present organization of documentation in East Germany are described. Originally, there was no coordination and all the work was done on local initiative. Now documentation work is coordinated by the Center for Scientific Literature established in 1950. Abstracting services and editors of other documentation literature were subordinated to that centre. Moreover, the Institute for Documentation was established at the German Academy of Sciences. There exists also the Institute for Agricultural Information and Documentation at the German Academy of Agriculture. The responsibilities of the institute include: documentation of domestic and foreign material, information, library research, theory of documentation, codification, personnel training, distribution of technical and scientific literature, etc. The article also describes the methods and the program of training documentation specialists and the existing facilities for such training. (C.B.L.)

10/65–89 **Information Exchange.** David F. Hersey. SCIE 147, 824 (1965). A letter discussing the activities of the Science Information Exchange (SIE) of the Smithsonian Institution. *(KG.)*

10/65–90 **The Number of Review Articles in Various Subject Areas of Chemistry.** Herbert J. Friedman, JCHD, v. 3, pp. 139-141, 1963.

A study of the twenty-five thousand items included in the first four volumes of *Bibliography of Chemical Reviews* reveals that the proportion of chemical review-type articles in applied subject areas is larger than the proportion in theoretical subject areas. At the same time, there is evidence that these 25,000 reviews provide over one million bibliographic references to the chemical and related scientific literature.

The publication, *Bibliography of Chemical Reviews,* is a product of the Chemical Abstracts Service and the American Chemical Society. Each volume of *Bibliography of Chemical Reviews (BCR)* collects together the abstracts of review articles included in one year of *Chemical Abstracts (CA).* The first four volumes of *BCR* cover the *CA* publication period of 1958-1961.

With the subject areas of chemistry defined in terms of the abstract classification system of *CA,* I found a great variation in the distribution of abstracts of review articles in the various *CA* sections. These percentages range from a low of about 3% to a high of over 28% of the published articles abstracted in the 33 major *CA* sections studied. Abstracts of patents are, of course, not included in these counts.

In this study, the subject areas of chemistry are stated in terms of the classification scheme used in *CA* through volume 55 (1961). However, in 1962, *CA* began using a new 73-section classification arrangement. Where possible in this paper, the equivalent section numbers in terms of the 1962 classification are also provided. (author)

10/65–91 **Scientific Information Exchange in Psychology.** William D. Garvey and Belver C. Griffith. SCIE 146, pp. 1655-9, (1964). Describes the immediate dissemination of research findings in the field of psychology. *(KG)*

10/65–92 **Cataloguing Code Revision: a Participant's Report on the Conference on Cataloguing Rules Held at Chaucer House in May, 1964,** P. A. Hoare. *Lib. Assn. Rec.,* 67 (1) January 1965. 10-12.

The Conference was called upon to discuss and comment on the effect of the draft rules drawn up by the British and American committees; there had been several alterations since the last similar conference in 1959 and the revisions could now be considered in the light of the 1961 Paris Conference and its seminal statement of principles. Rules were dealt with in groups and divergencies from A.A. Code rulings are described. Groups included: (i) Treatment of personal authors, (ii) Periodicals and serials, (iii) Form of name in heading, (iv) Corporate authors, (v) Subordinate bodies. Descriptive cataloguing was also discussed. (LSCA)

10/65–93 **Reprints: A Proposal.** Alan F. Hoffman, Jules Hirsch, et al SCIE 146, 1251 (1964). A letter suggesting that the distribution of reprints of scientific articles be handled by the publisher or by a third party rather than by the author, with a sufficient charge being made for reprints to cover the cost of the service. *(KG.)*

10/65–94 **Engineering Information Retrieval.**—B. E. Holm discusses information retrieval in the U. S., the philosophy of concept coordination, and the indexing system of value to A.I.Ch.E. members.—*Institution of Chemical Engineers, Transactions,* v. 42, no. 10; *Chemical Engineer,* no. 184, Dec. 1964, p. CE291 + 2 pages. (BATR)

10/65–95 **Maschinelle Dokumentation in der Elementarteilchenphysik** (Mechanical documentation in physics) Frederich Holtzhausen and Kurt Mellentin. NADO, 15 (3) September 1964, 130-133. References.

To meet the needs of quick information about physics research in institutes, a documentation unit dealing with high-energy physics has been established in the Deutsches Elektronen—Synchrotron (DESY), Hamburg. All accession to the DESY-Library will be examined by a physicist, who will choose items suitable for the proposed monthly publication entitled *High Energy Physics—and index.* Articles to be indexed are catalogued, allotted key subject words and then recorded on IBM punch cards. The monthly list (on offset folio) will contain the following parts: (i) Bibliographical section, giving title, author and source of articles indexed, with key word list forming an abstract as an annotation, (ii) Register of report numbers, listing the Institutes whose publications are being documented in the current issue, (iii) Register of authors, a straightforward alphabetical list of authors of the indexed articles, (iv) Key term register, the grouping of indexed articles under subject headings given in a key word thesaurus containing about 400 terms. The bibliographical section is to appear monthly, with quarterly cumulations of the registers of report numbers, authors and key words. *High Energy Physics* will cover yearly three thousand titles. Use of IBM punch cards for arrangement in different sequences enables the Register to be compiled at great speed, as ten thousand entries can be arranged in less than five minutes. This publication will appear in English. (LSCA)

10/65–96 **The Metals/Materials Division and ASM,** Marjorie R. Hyslop—SPLB, vol. 56, No. 3, pp. 185-187.

Cooperation and communication between technical and professional societies is a modern-day phenomenon that has evolved only during the past quarter-century. The relationship between Special Libraries Association, specifically the Metals/Materials Division, and the American Society for Metals is a prime example of the benefits to be derived from such communication and cooperation, even when the membership of the two organizations differs considerably in function, interests, and problems. To illustrate this evolution some history is briefly reviewed.

10/65–97 **The 600 Years of the Jagello Library,** Izsepy, E., *Magyar Konyvtari Szemle* (Hungarian Library Review), No. 4, 1964, pp. 366-368. In Hungarian.

The organization and development of the Polish University Library in Krakov is discussed. The library was founded in 1364 and was given the name of the Jagello University Library in 1400. Bibliographic work was started in the early 1800's. It has a rich collection of old books and manuscripts, although a large portion of holdings was lost during the Nazi occupation. However, since the end of World War II the holdings have increased to include over 1,000,000 books, 100,000 old manuscripts, over 100,000 old prints, 38,000 graphic illustrations, and 8,600 maps. The library publishes yearly lists of all foreign and domestic acquisitions, information bulletins and selected and special bibliographies. (DOC INC)

10/65–98 **International Cooperation of Scientific-Technical and Economic Information Centers in the Field of Agriculture,** Jiru, J., *Bulletin Technickych a Ekonomyckych Informaci* (Bulletin of Technical and Economic Information), No. 2, 1964, pp. 16-20. In Czech.

The international contacts and cooperation of the information centers of the Czechoslovakian Ministry of Agriculture, Forestry, and Water Resources are described. Czechoslovakia is represented in the International Committee for the Coordination of Scientific-Technical Information on Agriculture and Forestry of the COMECON nations. Problems of information exchange and cooperation were discussed at the International Conferences of Socialist Countries in Berlin in 1961 and in Prague in 1962. Bilateral agreements on exchange of information were signed between Czechoslovakia and other Socialist countries such as Poland, East Germany, and Bulgaria. In 1963 the Information Center of the Ministry of Agriculture was given the responsibility of coordinating the information activities of COMECON nations in the field of agriculture. The coordinating activites of that center, its publications, and plans for the future are reported. (DOC INC)

10/65–99 **The Network of Information Facilities in the Power Supply Industry,** Kadlechek, L., *Bulletin Technickych a Ekonomickych Informaci* (Bulletin of Technical and Economic Information), No. 2, 1964, pp. 8-15. In Czech.

The characteristic developments in the power supply industry are outlined. The specific requirements of this industrial branch regarding information and documentation services are investigated. To meet these requirements, the Czechoslovakian government established the Central Information Facility for Power Engineering and Supply. It has a staff of 42 employees, holds 120,000 volumes of books and periodicals in store, and processes 400 journals regularly. Information files are seached manually. The facility publishes the bulletin *Novinky Svetove Energetiky* (World's News of Power Engineering and Industry) and occasionally other reference and abstract services. The structure of the facility, cooperation with government agencies and research institutes in Czechoslovakia and abroad, and other aspects of its activities are described. (DOC INC)

10/65–100 **A Single Publication Group for the ACS.** Richard L. Kenyon, CENE *42*, 7 (December 28, 1964). The American Chemical Society will now have a single, unified organization for the editorial development and production of its primary publications. The main goal will be simple: constant improvement of chemical and chemical engineering publications to meet the needs of scientific and technical people. (HLW/ABI)

10/65–101 **Documentation and Information, a Matter of Importance for the Federal Republic of Germany.**—A detailed survey of the state of documentation in Western Germany is made. (In German.) Dokumentation und Information als Angliegen der Bundesrepublik Deutschland,—Heinz Lechmann. NADO v. 15, no. 4, Dec. 1964, pp. 157-166. (BATR)

10/65–102 **The National Referral Center: Science and Technology in the Library of Congress,** M. W. McFarland, ASLP, 16 (8) August 1964, pp. 258-268. References.

A brief historical introduction emphasizes that the Library is for, and is responsible to, Congress, The structure of the Library is outlined and the responsibilities of each section concerned with here, science and technology are described. A diagram shows the function and organization of the Science and Technology Division; the work of this division is discussed in some detail. The Referral Centre went into operation in March 1963 and is supported by the National Science Foundation, but it is administered and operated entirely by the Reference Department in the Library of Congress. The functions and activities of the Centre are described in detail. A recent analysis of users showed that 51% of inquiries came from commercial organizations; 18% from Government agencies, including libraries; 8% from educational institutions; 5% from professional societies, technical and trade organizations; and 18% undisclosed by the survey. Anyone can use the Centre, the services of which are free. It is hoped that the Centre's lead will encourage the development of other "referral centres" throughout the world to provide scientists and engineers with ready access to scientific and technical information everywhere. (LSCA)

10/65–103 **Tests on abstracts journals,** John Martyn and Margaret Slater. IDOC 20(4) December 1964, pp. 212-235.

Report of an investigation by the Aslib Research Department into the various losses of information which occur in the use of abstracts journals. Loss of information can occur in three ways: (i) An item in a journal may not be abstracted at all; (ii) When using the subject index of an abstracts journal, all the relevant items which are abstracted are not indexed under the terms by which the search is defined; (iii) Of the references, located via the subject index, not all are available in British libraries. The method adopted in the investigation was to select a recent comprehensive bibliography of a particular subject as representing the total volume of published literature on the subject during the period which it covers; to check the author and subject indexes of the appropriate abstracts journals to determine their coverage; and then to determine how many of the references are available. Results are given in tabulated form for twelve subjects, and some general conclusions are drawn from these results. (LSCA)

10/65–104 **The Organisation of Science Information Activity in the Soviet Union,** A. I. Mikhailov, REDO, 31(4) November 1964. 143–148. (In Russian with summary in English.)

The national information network is headed by the State Committee for co-ordination of scientific research of the USSR. It includes, in addition to the four central institutes, general information institutes in each republic, central specialized information and research institutes in each sector of activity, informaton bureaux in each economic region, and information departments in factories and institutions. VINITI was founded in 1952 and concentrates on abstracting in the science and technology fields. In 1963 it received 13,500 periodicals from over 100 countries. The abstracting service includes 157 titles and one third of the subscribers are foreign. It also published 67 express-information series and index cards. The library contains 1m. items. Various forms of co-operation exist with international bodies, e.g. exchange arrangements, conferences. Post-graduate training — three years full-time or four years part-time—is organised. Candidates must present a thesis in one of these fields: information handling as a whole; material information carriers; equipment and methods of storage and retrieval. Organisation of optimal information service is regarded as an important government task. Efforts are now being concentrated on mechanization and automation of information work. (LSCA)

10/65–105 **Citation Index und Ruckwartskatalogisierung Beispiele fur Zitatendokumentation.** (The *Citation Index* and the *Ruckwartskatalogisierung* as examples of citation documentation). Fritz Model. NADO 15 (3) September 1964, pp. 122-130. References.

A short historical survey of citation documentation using as illustrations *Shepard's Citations; Science Citation Index, Code of Citations* and others. Particular comparision is made between the *Science Citation Index* and the *Ruckwartskatalogisierung* (*i.e.* Hydrographic documentation of the German Hydrographic Institute). The scope of the two indexes varies; the *Science Citation Index* includes all articles on life and physical science, which appear in a given small number of key journals. The Hydrographic Institute's catalogue covers every hydrographic article, from a survey of nearly 2,000 journals. In addition references are kept for 10 years only in the *Ruckwartskatalogisierung* to keep information really current. (LSCA)

10/65–106 **Private and Public Communications in Physics,** Michael J. Moravcsik. *Physics Today,* March 1965, pp. 23–26.

Almost two years have passed since the Weinberg panel issued its report. It was a thorough and well-considered study, containing a large number of recommendations. I am not sufficiently familiar with the detailed organization of the various sciences to be able to tell whether, on the whole, the Weinberg report's recommendations have been realized or not. I do know, however, that in that sector of physics which I have contact with and among the physicists I know, its suggestions have, so far, had a negligible impact. In a sense, this is not surprising, since a comprehensive study like the Weinberg report has, in addition to its many advantages, at least two drawbacks. First, by its very nature, its recommendations involve new organizations or new programs on a large scale, which have to be established by the government or by large private institutions, and, as such, require time, money, and initiative which are not always present. Second, again by the same aspect of its nature, its recommendations appear, at first sight, to be unsuited to any action an individual scientist, or a small group of scientists, can undertake on their own, and hence do not foster the feeling of individual responsibility and participation which the Weinberg report, in the abstract sense, stressed very emphatically.

The purpose of the present article is, therefore, to make a few very specific suggestions concerning the communication problem in physics. They are, at best, partial realizations of the recommendations of the Weinberg panel, and can be considered as stopgap measures until permanent remedy arrives through the large-scale reorganization of our scientific communication habits. Neither are they a complete answer to the problems involved. Rather, they are simply a collection of ideas, which might either be tested in practice, or, if they prove to be too "outrageous," will, I hope, be replaced by better ideas by others. Their advantage lies in the fact that they are all easily realizable by a small group of physicists in a short time with essentially no investment of money. I believe that, just as in foreign aid, there is room, beside the large-scale development projects, for local "self-aid" projects furnishing quick, partial relief to specific bothersome problems.

10/65–107 **The Role of the National Library in Science and Technology with Special Reference to the United States Library of Congress,** L. Quincy Mumford. UNDL, 18 (4) July–August 1964, pp. 172–177, 192.

The role of libraries in the scientific and technological complex lies in improving the effectiveness of communication among the scientists and engineers who can apply technical information. In planning the national library regard must be had to various national factors (*e.g.* pattern of government), the extent to which centralization is desirable and the balance between professional and (in this case) scientific competence. The collections of the library will depend upon the needs of the particular nation, but should include a complete collection of the country's own output of scientific documents and an authoritative representation of fundamental as well as relevant monographs, serials and textbooks. The importance of acquisition by exchange is stressed. The library's approach to providing services must be dynamic. The remainder of the article discusses four areas of service particularly important for a national library with responsibilities in science and technology. (i) Referral service; a short account of LC's National Referral Centre for Science and Technology is given. (ii) Reference and bibliography services, which includes an account of LC's Science and Technology Division. (iii) Services to the government, which describes LC's Legislative Reference Service, and (iv) Co-operation with other libraries. (LSCA)

10/65–108 **LC's International Organizations Section,** Kathrine O. Murra. REDO, 31 (4) November 1964, pp. 138–142. Diagr., references.

Brief history and description of operations. Information exchange agreements exist with 2,200 international organizations. A reference service has been developed and since 1959 the Section has published the quarterly World List of Future International Meetings. In 1963 the Section issued International Scientific Organizations, a guide to their libraries and services. (LSCA)

10/65–109 **Input and Output—Technical Information Retrieval —Lists Information Services for the Plastic Industry.** Describes different types of abstracts, coordinate indexing system with roles, activities of a technical information center, mechanization, and the technical journal and information system of the future. Louis I. Naturman, *SPE Journal,* v. 21, Jan. 1965, pp. 17–32. (BATR)

10/65–110 **The Czechoslovak Institute for Scientific and Technical Information as the International Center for the Coordination of Scientific and Technical Information in Agriculture for Member States of the Council of Mutual Economic Aid.** Opava, J. *Technicka Knihovna* (Technical Library), No. 3, 1964, pp. 79–82. In Czech.

The author surveys the functions of the Institute of Scientific and Technical Information, established in 1961, in its new role

as the international center for agricultural information. The institute publishes considerable material on agriculture, and its agricultural library is now the 4th largest in the world. The library has 470,000 publications in stock and regularly receives 2,300 agricultural periodicals from 70 countries. It exchanges information with 1,051 foreign sicentific institutions in 68 countries. The services of the institute are available to the agricultural organization of all the member nations of the Council of Mutual Economic Aid. First class experts evaluate all pertinent data and make recommendations. The institute publishes scientific and popular scientific periodicals, bibliographies, reports, and monographs on selected subjects. Coordination of cataloguing, processing, and bibliographical work of the information centers and research organizations of the member nations should provide the basis for quick and effective documentation service. (M.H.G.)

10/65–111 **The Biomedical Information Complex Viewed as a System.** Richard H. Orr, Gregory Abdian, Charles P. Bourne, Edwin B. Coyl and Alice A. Leeds. Federation of American Societies for Experimental Biology, Washington, D. C. 1964, 13p. Grants PH43 62 167; GM09166. Unclassified report. Reprint from Federation Proceedings, 23:5, pp. 1133-1145, Sept.-Oct. 64. (Cpies not supplied by DDC) AD–454 269 Div. 16.

To aid in visualizing and understanding the heterogeneous aggregate of interdependent operations, activities, and services that handle the information generated by, and used in, biomedical research, this complex was analyzed as a system from a viewpoint of the functions it performs. The result was a qualitative model with the following major functional components: (1) generation and use, (2) oral communication, (3) recording and distribution. (4) document processing, (5) information processing, and (6) control. Bewteen generation and use, the flow of information through components (2), (3), (4), and (5) depends upon parallel and sequential chains of processing operations. The operations of each component depend, in general, upon the prior accomplishment of the operations of the preceding component. The capacity of a given component is limited to that of its slowest operation except where alternative paths exist. The costs of operating this complex are met by government, private foundations, industry, academic institutions, and user fees for services (such as subscription fees). The present trend is toward increasing dependence on government support.

10/65–112 **Biomedical Literature: Volume, Growth, and other Characteristics.** Richard H. Orr, and Alice A. Leeds. FEPR vol. 23, pt. 1, pp. 1310–1331, 1964.

This study's aim was to gather and analyze reliable data on the quantifiable characteristics of the biomedical literature that affect communication problems within the research community. Previous studies and standard bibliographic compilations were critically reviewed, particularly for data that could be used to determine changes with time. In addition, all publications generated during 1961-1962 by the extramural and intramural research programs of the National Institutes of Health (NIH) were analyzed as samples of the current document output of U.S. research. The world's substantive biomedical serials numbered about 5,700 in 1960, an increase of less than 7.5% since 1950; and during this decade, the best available evidence indicates that the total number of papers published annually in these serials increased by around 20%. This rate of growth means a doubling of the serial literature in 38 years. The proportion of the world's biomedical serials published in the U.S. remained constant (about one-fifth) from 1950 to 1960, and gross language patterns were relatively stable for the literature as a whole. Approximately 500 technical books in biomedicine are currently being produced each year in the U.S., and the total accumulation of biomedical monographs is doubling about every 32 years. Roughly 2,000 U.S. technical reports on biomedical subjects are being issued annually; only a few are security-classified, but many of the remainder are, for practical purposes, inaccessible to scientists who are not eligible for the special services that have been set up by governmental agencies. Of 15,979 documents generated in 1961-1962 by grantees of NIH, 90% were published in journals; some 100 "core" journals contained two-thirds of all grantee publications and 10 accounted for over one-quarter of all the publications in journals. Analyses of the distribution of sample documents among different journals and books show that the biomedical literature is more scattered than that of chemistry or physics, but the degree of scattering varies widely from one subfield to another. The findings concerning languages and scatter have important implications for biomedical research workers and for information services, such as libraries and abstracting-indexing services. Concern over the "literature explosion," the statistics of which seem to have been based on inappropriate data, has led to an overemphasis on volume as a factor in the communication problems of biomedical scientists and in the difficulties of information services attempting to serve their needs. (authors)

10/65–113 **Document Retrieval: The National Biomedical Library System and Interlibrary Loans.** Richard H. Orr and Vern M. Pings. FEPR, vol. 23, pt. 1. pp. 1155–63, 1964.

By 1965 improved reference retrieval services, such as those to be provided by the Medical Literature Analysis and Retrieval System (MEDLARS) of the National Library of Medicine (NLM), will enable biomedical scientists to obtain references to relevant documents more easily, and from a broader segment of the world's scientific literature, than at present. References to relevant materials are, however, of no value to the scientist unless he can obtain the documents referred to (document retrieval). The aim of this study was to explore the likely impact of improved reference retrieval services upon the present "system" that supplies the biomedical scientist with the documents he requests, i.e., the biomedical library "system." By reviewing past studies, considering available evidence ,and collecting such new data as time and resources permitted, the document retrieval operations of this system were analyzed. From their own collections, biomedical libraries supply varying percentages of the total documents requested by their respective local biomedical communities. They meet the remaining requests by calling on other libraries for interlibrary loans. The volume of these loans has been increasing by about 10% a year and presently exceed 500,000 documents annually. The cost of maintaining this flow is over $2 million per year. Current signs of strain indicate that the system, as presently operated and financed, has reached its maximal capacity to meet the demand for interlibrary loans and is critically unstable. Present evidence indicates that this demand may reach one million documents annually by 1965. The capacity of the system must be increased rapidly if biomedical scientists are to benefit from the new and improved reference retrieval services that will soon be available. There are three major alternatives for a long-term program to meet this challenge: *1)* increase the capacity of NLM to furnish photocopies to libraries on request, *2)* establish regional loan centers, and *3)* develop the collections of local biomedical libraries to reduce the need for interlibrary loans. As an interim measure, subsidy of interlibrary loan operations is suggested. (authors)

10/65–114 **Generation of Information: Published Output of U.S. Biomedical Research.** Richard H. Orr, Gregory Abdian, and Alice A. Leeds. FEPR, vol. 23, pt. 1, pp. 1297–1309, 1964.

The aim of this study was to determine quantitative relations between the magnitude of the U.S. biomedical research effort and the volume of published documents it generates. Statistics on the funds, manpower, and organization of this research effort over the past two decades were compared with data on the number of documents that could be attributed directly to work supported by the major sources of research funds. The document output of U.S. research was also assessed indirectly by collecting and analyzing data on the publication habits of biomedical scientists. During the period 1957 to 1961, the ratio of the number of papers generated each year by the extramural program of the National Institutes of Health to the number of research grants active two years earlier was relatively constant. In contrast, the ratio of research expenditures to document output in this and other major programs increased steadily. Serial data on the publication habits of biomedical scientists in specific institutions show that, as individuals, they are writing neither more nor less papers now than previously; but a steady, general trend toward multiple authorship suggests tentatively that the over-all ratio of the number of papers published annually by these populations to the number of scientists may be decreasing. The total document output of U.S. biomedical research in 1960 is estimated at 26,000

papers, or about half the total number of U.S. biomedical papers published in that year. Over the past decade, document output has increased at roughly the same rate as research manpower and approximately one-half as rapidly as total research expenditures. These findings have implications for the users of biomedical literature, for the information services that must handle the document output of biomedical research, and for those concerned with the development of the national research effort. (authors)

10–65/115 **Reference Retrieval Tools: Biomedical Abstracting and Indexing Services.** Richard H. Orr, Vern M. Pings and Alice A. Leeds. FEPR, vol. 23, pt. 1, pp. 1164–76. 1964.

The primary objectives of this study were: *1)* to analyze the general functions of abstracting-indexing services; *2)* to describe quantitatively the operations, products, and performance of services useful to biomedical research workers; and *3)* to identify the problems associated with maintaining and improving the performance of U.S. services covering biomedical literature. Scientists use the bibliographic tools produced by these services for three general purposes: *1)* alerting, *2)* searching, and *3)* informing. The criteria by which a service's performance is judged differ for each of these use-functions, and compromises are necessary when a single tool must serve more than one purpose. In the past decade, mounting document loads, processing costs, and personal shortages, coupled with demands for increased performance, have resulted in four major trends in U.S. services: increased mechanization, more single-purpose bibliographic tools, increasingly narrow specialization in subject scope, and more cooperation among services. Three hundred twenty-six foreign and 147 U.S. services process biomedical literature. Altogether these services process almost two million documents yearly; but all these documents are not biomedical, and this figure includes many biomedical documents processed by more than one service. In the past 10 years, the combined output of the larger U.S. services has increased much more rapidly than the biomedical literature has grown. The present study assessed the coverage by English-language services of a large sample of the documents currently being produced by U.S. biomedical research workers. Of the 891 journals in which the sample documents were published, all but 39 (96%) were "covered" by one or more of six major *Index Medicus (IM)*, *Chemical Abstracts (CA)*, *Biological Abstracts (BA)*, *Excerpta Medica (EM)*, *Bibliography of Agriculture*, and *Psychological Abstracts (PA)*. Duplication of coverage was considerable; the sample journals were, on the average, covered by three of the six services. *IM* covered 66% of thse journals and processed 87% of the sample documents; however, because of the selective policies of *CA*, *BA*, *EM*, and *PA*, it is likely that a significant percentage of the sample documents were not abstracted by any of these major services. Abstracting and indexing services useful to biomedical scientists have improved significantly in the past 10 years with regard to completeness of coverage; however, with rising unit costs and the demand for better performance, closer cooperation among the services and coordination of their efforts are imperative. (authors)

10/65–116 **Trends in Oral Communication Among Biomedical Scientists: Meetings and Travel.** Richard H. Orr, Edwin B. Coyl, and Alice A. Leeds. FEPR, vol. 23, pt. 1, pp. 1146–1154, 1964.

This study was undertaken to provide factual bases for a broad consideration of communication problems in biomedical research and focused primarily on the formal type of oral communication, i.e., exchange of information at meetings. Data were collected on the growth in number and size of biomedical meetings, on services announcing forthcoming meetings, on international exchange of oral information, on scientists' travel, and on government support of biomedical meetings. In 1961, there were almost 500 U.S. biomedical societies, which held 1,500 regular meetings. The number of such regular meetings has trebled in the past three decades. Meetings that serve large segments of the biomedical research community have, since 1957, grown by 10–20% annually. In the past 10 years, the number of U.S. personnel working abroad in biomedical fields has increased 50%; but for other fields, the increase has been much larger. Of funds provided by the National Institute of Health for direct support of all types of information activities, over 25% has gone to meetings. The importance of unanswered questions and the large amount of scientists' time devoted to oral communication merit more study than this mode of scientific communication has previously received and warrant increased efforts to improve both meetings and informal oral communications. (authors)

10/65–117 **Unesco and International Co-operation in the Field of Scientific and Technical Documentation,** A. Perez-Vitoria. UNDL, 18 (5). September-October 1964, pp. 207–210. References

A brief account of Unesco's work in this field under four headings: (i) Co-operation with member states; (ii) Co-operation with non-government organizations; (iii) Publications, and (iv) Unesco's long-range programme in scientific documentation. Section (iv) lists the recommendations of three working parties appointed by Unesco between September 1963 and January 1964 to consider respectively problems relating to scientific publications, automatic documentation (storage and retrieval) and scientific translation and terminology. Most of these recommendations will probably be included in Unesco's programme for 1965–66. (LSCA)

10/65–118 **Unesco's Contribution to the Development of Scientific and Technical Documentation Centres** [part 1], A. Perez-Vitoria, UNDL 19 (1) January–February 1965, pp. 2–23.

Forty specialists from twenty-one countries have taken part in initiating the development of a world-wide network of centres within the framework of the United Nations Technical Assistance Programme. Twelve centres have already been established and this text reviews what has been accomplished and shows what problems have been solved and what remains to be done under the following headings: subject field, holdings, periodicals, library and reading room, publications, document reproduction unit, bibliographic service, translation service, miscellaneous activities and publicity. The question of whether services should be free or paying is also discussed. (LSCA)

10/65–119 **Technical–Scientific Information Activities of the Union of Technical and Scientific Societies,** Philip, M., *Tudomanyos as Muszaki Tajekoztatas* (Scientific and Technical Information), Special Issue, Dec. 1963, pp. 115–138. In Hungarian.

Twenty-five scientific associations are members of the Union of Technical and Scientific Societies. These scientific associations organize lectures, symposia, conferences, debates, exhibitions, technical film presentations, professional post-graduate courses, and professional consultation services covering the most important fields of industry, agriculture and natural science. The union as such deals with problems which affect all the member societies. The member associations publish 45 professional papers with a total circulation of approximately one million. The Home of Engineering (i.e., the Budapest Center of the Union) is one of the most significant establishments for technical-scientific information. A complete list of the member societies of the union is included. (author)

10/65–120 **Reprint System Debated.** George V. Pickwell and Everett Douglas, John A. Blazer, J. Sri Ram, Michael R. Cummings, Lester Goldstein, D. H. Hubel, et al.; Frank T. Manheim. *SCIE 147*, 677, 679 (1965). A group of letters discussing suggestions on reprint distribution made by Hofmann, et al. (*K.G.*)

10/65–121 **The Scientific Journal—300th Anniversary**—J. R. Porter. *Bacteriological Reviews*, v. 28, no. 3, Sept. 1964, p. 211–230. (BATR)

10/65–122 **Activities of Special Committees,** Riepert, R., and Selbmann, E., *Mitteilungen aus dem Wissenschaftlichen Bibliothekswesen der Deutschen Demokratischen Republik* (Reports on Scientific Libraries of the German Democratic Republic), No. 5, 1964, p. 56. In German.

Activities of the librarians' Committee on Statistics and Committee on Circulation are discussed. The Committee on Statistics met in March 1964 in the library of the Dresden Technical University. Topics discussed were: principles, types, and methods of surveys of library statistics in Socialist countries; basic rules for collecting and evaluating statistical data on scientific libraries; and the preliminary design of a statistical data-sheet for all scientific libraries that have a full-time administrative staff. The Com-

mittee on Circulation worked on the draft-proposal for new library loan procedures. (DOC INC)

10/65–123 **A Calculus for Journal Publishers.** Allan M. Russell. *SCIE 147,* 110 (1965). Comments on Wolfe's editorial "Basic research journals." (*K.G.*) (See 10/65–132)

10/65–124 **Les travaux de l'Union des Associations Internationales dans le domaine de la documentation internationale.** Speeckaert, George Patrick. REDO, vol. 32 no. 2, May 1965, pp. 56–58.

The principal aim of the UIA is to establish and disseminate a basic, accurate and complete record of the channels of international cooperation, in particular international organizations, meetings and publications.

As reference tools, the UIA publishes the Yearbook of International Organizations, the International Congress Calendar, current and annual Bibliographies of international meetings and a selective Bibliography on international organization.

The UIA attempts furthermore to further the documentation and information activities of international organisations in their particular fields of interest. (REDO)

10/65–125 **Methods and Organzation of Information Work,** Schmoll, G., DKMN, No. 3, 1964, pp. 78–81. In German.

The author discusses the importance of choosing the right system for development of an unified information service system in East Germany. Prior to actual implementation, a certain amount of basic research in the design of information systems is required. The purpose of the information service is to promote economic progress; however, it is necessary to introduce new methods and ideas in the information service. Some methodological problems to be solved are the evaluation of working operations, organization of work, tools and means of work, choice of equipment, and application of mechanized devices. It is also imperative to employ qualified personnel and to utilize available information sources. (DOC INC)

10/65–126 **The New York Public Library,** Marion L. Simmons. *Bookmark* (N.Y. State Library), 24(2) November 1964, 35–42. Illus.

The library system owes its origin to a gift of over $5 million for buildings from Andrew Carnegie in 1901. 39 of the present 80 branches were built from this fund. The library has over 7m. v. (over 3m. v. for circulation), ¾m. pamphlets, 9m. mss., 150,000 recordings, 135,000 pieces of sheet music, 125,000 prints, 60,000 reels of microfilm, 265,000 maps, 2m. mounted pictures, 800 sets of periodicals. One million people are members and circulation in 1963 was almost 14m. Due to staff shortages, only the thirty largest branches are open on Saturdays—yet there are 2,700 on the staff, 661 being professionals and 60-65 trainees. Research materials (in over 3,000 languages and dialects) are found in the Central Research Library on Fifth Avenue. This is visited by more than 8,500 people every day, with a record day each year of about 20,000. This library operates on private funds. In the catalogue there are 10m. cards; those on the First World War occupy 73 drawers. Some parts of it have been photographed and turned into book catalogues. Two projects are in progress. In 1965 the Library and Museum of the Performing Arts will be opened at Lincoln Center, the city's new cultural centre. This will bring together the Music, Dance, and Theatre Collections, while a circulation collection of books, records and films is being assembled. In 1966 a new central library will be opened on Fifth Avenue and 40th Street adjacent to the present building. It will have have 300,000v. on open shelves to provide reference and lending materials mainly for college students and the general reader. There will be seats for 1,000 readers. This will provide a self-service for people not adequately provided for at present in mid-town. It will relieve the pressure on the older building and make it possible to give a better service to professional, business and industrial researchers. (LSCA)

10/65–127 **Technical Documentation Center of Forestry,** Stefan, M., *Calauza Bibliotecarului* (Library Guide), No. 1, 1964, pp. 49–52. In Rumanian.

The Department of Forestry organized this center in 1960 with the purpose of advancing the forestry science in Rumania. The center has a library, bibliography and documentation service, and an information service. It organizes conferences and publishes monographs and periodicals. The library has 33,000 volumes and exchanges documentary material with 29 countries. There are available about 1,200 titles of domestic and foreign journals. The material is being indexed according to the Oxford Decimal Classification system for forestry science. A single documentalist is responsible for the processing of an average of 30 to 40 foreign magazines. One of the publications of the center is the yearly *Rumanian Bibliography of Forestry*. Other periodicals are: *The Bibliographical Index*, 11 volumes annually, each containing approximately 400 titles of foreign and domestic articles, and *Current Documentation on Transportation*, 12 volumes per year of about 40 pages each. A multilingual dictionary of forestry in Rumanian, Russian, Hungarian, German, and French is being prepared. (F.S.)

10/65–128 **Founding of the German Library Association,** Unger, W., *Mitteilungen aus dem Wissentschaftlichen Bibliothekswesen der Deutschen Domokratischen Republik* (Reports on the Scientific Libraries of the German Democratic Republic), No. 4, 1964, pp. 37–43. In German.

The German Library Association was founded in March 1964 in East Berlin. The meetings were attended by 110 librarians of the German Democratic Republic, representatives of the central government and party, and delegates from library associations of various Socialist countries. At these meetings the constitution was adopted and members of the executive committee were elected. The article lists the names, occupations, and functions of the members of the executive committee; the locations of the regional committees as well as the names and occupations of their appointees; and the names of the appointees to special working committees. The address of the association's newly opened office is Georgenkirchstrasse 24, Berlin C 2. (DOC INC)

10/65–129 **Second Thoughts on Scientific Information.** Weinberg, Alvin M. CRLI, vol. 25, no. 6, November 1964, pp. 403–71.

The author is concerned here with what he considered to be one of the main messages of the Weinberg Report (*Science, government, and information*) that science in response to the information crisis is undergoing a hierarchial social reorganization which will impose a corresponding reorganization on the scientific information system.

Professor Wigner, a member of the Panel responsible for the Report thinks that the scientific community is layering itself into a hierarchy in which at the lowest level are the bench scientists, each working in a narrow field and communicating with other bench scientists working in closely related fields and at the next level are the group leaders who communicate with one another and thus form a contact between different groups of bench scientists. Supergroup leaders may be above the group leaders who communicate with one another at a higher level of abstraction and maintain a higher order surveyance over the scientific community.

In sciences, such as physics, which have a well-defined theoretical structure, the job of higher order surveillance has been taken on by the theoretical scientists. They interpret and correlate the bench scientists' results to try to make generalizations encompassing all data. In the less well-defined sciences, such as chemistry and biology, the information centre may carry out some of this work. These centres must have all relevant data and be manned by scientists who can interpret the data and understand the documents held. They should be run by recognized scientific leaders in the subject fields of the centres. They are technical institutes rather than technical libraries. The librarian could, however, contribute to their successful operation by helping to develop and support systems which will give a 99.5 per cent retrieval level in searches, and also by devising systems for quickly finding what information has been compiled and where to get a copy of the compilation. The centres should be located in large laboratories or in universities.

The Government has acted on many of the suggestions in the Report. Information has become the concern of government agencies whereas previously only libraries and information specialists were concerned with it. The idea of the specialized information centre has caught on, but the idea of a delegated agency being responsible for co-ordinating and supporting information in a field which is the responsibility of several agencies has not done so because of political difficulties. The information centre

may be able to do at least part of the work proposed for the delegated agency. Paper prepared for presentation before the Association of College and Research Libraries and the ALA, St. Louis, Missouri, 30 June 1964. (JDOC)

10/65–130 **Progress Report on the Book Processing Centre,** Clara Wendel. *Florida Libs., 14* (3) September 1963, pp. 7–10, 32.

By 1961, many public librarians in Florida considered it desirable to have a joint book processing centre, and one was established at the public library in Orlando in 1962. Financial aid comes from the Florida State Library which advances money from the rural library development fund, this being made available to the State by the Federal Government under the Library Services Act. The grant makes good the difference between the cost per book to libraries of 75 cents and the working costs of 88.5 cents. The annual output is about 40,000 books, staff the equivalent of ten full-time members, and libraries of twenty-three counties are represented. Cards are printed by IBM typewriters and A. B. Dick offset press, though later it is hoped to introduce Xerox equipment for copying LC proof sheets. Multiple-copy order forms are sent by libraries, and these titles are checked at the centre against spare cards from previous orders and against LC sheets. A considerable proportion of books are thus already catalogued on arrival; others await LC entries or are catalogued on the spot. Typed labels and plastic jackets are used on all books with dust jackets, others are hand lettered or labelled. Problems include the failure of many libraries to order promptly at regular intervals, and delays in the receipt of funds, both of which cause back-logs in ordering. (LSCA)

10/65–131 **How to Improve Information Work,** Winde, B., DKMN No. 3, 1964, pp. 65-67. In German.

Information facilities in East Germany should be better utilized. This depends to a great extent on the cooperation of various executives in industry and administration. First, the information needs and capabilities of their enterprises, viz. state-operated factories and plants, must be evaluated. The Central Institutes of Information and Documentation has the responsibility of developing a uniform information system and coordinating such activities. Information centers, which hitherto operated mostly on their own, should start cooperating closely with other comparable institutions. Information specialists should prepare state-of-the-art studies on topics in which industry is interested. (DOC INC).

10/65–132 **Basic Research Journals.** Dael Wolfle. *SCIE 146,* 869 (1964). Discusses the financial situation of basic research journals, and suggests that a reduction in their number would lead to merged journals that are stronger and healthier. *(KG)* (See to merged journals that are stronger and healthier. (KG) (See 10/65–123).

10/65–32 **Patentinformation** (Patent information). Bengt R. Wurm. *Tid.f. Dok., 20* (6) 1964, pp. 73–83.

The patent is a technical description at the same time as it defines an exclusive right. This places special demands on the Patents Office in their search for and retrieval of information in conjunction with their examination of novelty and patentability. Information retrieval systems must permit search both on the specific and general levels. The stock of letters patent used in international search numbers some 7-10m. and increasing at the rate of 200,000 a year. Attempts to mechanize the examination of novelty during the Post-War period have now been co-ordinated through ICIREPAT (International Co-operation in Information Retrieval among examining Patent Offices). (f. 1961). An inventory showed that about 80 systems, based on edge-notched, peek-a-boo and punched cards and computers, had been devised or were under development. A research Associates Programme started in 1962.

10/65–133 **A National Library of Science System Has Been Proposed.** Dr. Stafford L. Warren. *CENE 42,* p. 21 (June 8, 1964).

Dr. Warren outlined his plan to a Conference on Automation and Library Planning at the University of Pittsburgh. The plan calls for creation of a computer-based literature pool. These data, supplied on tape and microfilm, would be contributed by federal and other libraries or contractors from their own or assigned holdings in the scientific literature. The total literature pool thus developed would be replicated and used to set up seven or more regional centers. (ABJ/HLW)

10/65–134 **Pitt Attacks Information Retrieval Problem.** *CENE 42,* p. 46 (June 8, 1964).

University of Pittsburgh's new library and information sciences program will develop knowledge availability systems. To this end, Pitt established a new school called the Graduate School of Library and Information Sciences. The University also plans a comprehensive attack on the problem of keeping scientists and educators abreast of the development in their fields. The library will contain special areas for experimentation in the most advanced technology of information handling. (ABJ/HLW).

10/65–135 **Organization of Scientific Information in the Polish Academy of Science.** Wysocki, A., *Aktualne Problemy Informacji i Dokumentacji* (Current Problems of Information and Documentation), No. 1, 1964, pp. 18–21. In Polish.

The activities of the Polish Academy of Sciences include research in the information field, training of information specialists, development of technical information facilites, and coordination of scientific information within the academy itself. The Documentation and Information Center of the Academy is responsible for the implementation of this program. The functions of that institution are described. They include coordinating information and documentation services of seventy-eight institutes and agencies of the academy. The Polish Academy of Sciences publishes a number of information and documentation journals and monographs, among them two English language periodicals: *Quarterly Review of Scientific Publications* and *Polish Scientific Periodicals—Current Contents with Author Directory.* (DOC INC)

10/65–136 **Auerbach Reorganizes to Form Three Divisions.** *CENE 42,* p. 21 (June 8, 1964).

The three divisions are: Information Sciences Division; Information Products Division and an International Division. Auerbach, primarily an information systems design and consulting organization in Philadelphia, Pa., has also set up a new program-development department, on the staff level. The department's responsibility is to adapt the company's techniques of handling and using information to specific industries. (ABJ/HLW)

10/65–136 **Better Information Coordination Urged.** *CENE 42,* p. 27 (November 30, 1964).

House committee says there should be a central group capable of enforcing agency cooperation. This statement is the result of a study of documentation and dissemination of research and development results. The committee recommends that a single clearing house be designated to coordinate all foreign federal activities in documentation and dissemination of scientific and technical information. It also calls for frequent review of security information so that such material is not kept under wraps any longer than necessary. This report raises more questions than it answers. (HLW/ABJ)

10/65–137 **Common Patents for the Common Market.** *CENE 42,* pp. 86–106 (June 15, 1964).

A single patent that covers more than one country, a long discussed concept, may finally become a reality. European experts work diligently to develop a unified EEC patent system—to stimulate trade, cut out duplication of effort and simplify patent procedures. (HLW)

10/65–138 **Computer Searching of U. S. Patents** *CENE 43,* 31 (April 26, 1965).

Information for Industry, Inc. and Gulf Research and Development Co. have a license agreement making available new developments in computer searching of U. S. chemical patents. Gulf has revised the information contained on IBM 1401 magnetic tapes and made them suitable for use in an IBM 7094. In addition, it has prepared a frequency distribution table of descriptive indexing terms. (HLW)

10/65–139 **From Original Journal to Abstract — 90 Days** *CENE 43,* pp. 76–78 (June 14, 1965).

New Chemical Abstracts Service building is geared for

speed, efficiency and expansion. The new five-floor building will take care of the current need of CAS and its expected expansion for a few years. Many new publications and services are in various stages of development at CAS. (HLW)

10/65–140 **J. Phys. Chem. Editor Outlines Policies** *CENE 42*, pp. 57, 64 (December 21, 1964).

Dr. Frederick T. Wall announces that new approaches to the problem of scientific communication will be examined carefully in the months ahead. Theoretical and experimental articles, notes, and communications are acceptable, provided they represent distinct contributions to the literature. Notes must be brief and concise and should report definite results. Communications should deal with timely research items of unusual importance. Symposium papers may be published as a group, but only through special arrangements with the editors. (HLW/ABJ)

10/65–141 **Microfiche Standards Adopted** *CENE 42*, pp. 54-55 (August 31, 1964).

Three government agencies have adopted standards to speed up and simplify dissemination of R and D reports. The three agencies—Atomic Energy Commission, National Aeronautics and Space Administration and Department of Defense—will each have switched over to the new system by the early autumn. In operation microfiche (negative cards containing documents in micro form) will be used to reproduce all R and D reports from the agencies and their contractors. (HLW)

10/65–142 **Publications Structure Revamped** *CENE 42*, pp. 53-54, 64 (December 21, 1964).

This action puts all publishing operations except Chemical Abstracts Service under one administrative head. Beginning January 1, 1965, the American Chemical Society will eliminate the present administrative division between the Applied Journals and the Fundamental Journals. Dr. Richard L. Kenyon will become Director of Publications in the combined organization. He will coordinate the broad editorial policies and plans, advertising, and circulation programs of the Society's publications. (HLW/ABJ)

10/65–143 **Science Library System Gets Mixed Reaction.** *CENE 42*, pp. 21-22 (June 15, 1964).

A proposal by Dr. S. L. Warren to establish a National Library of Science System could have far-reaching effects on the journal system of reporting scientific research. However, the system's backers allay fears that pooling published journal literature would mean the end of journals. The scheme carries with it a suggested six-year price tag of $308 million. First step in the plan is to tie in and augment existing scientific information programs into the one system. (HLW)

Reports of Facilities in Operation—Libraries

10/65–144 **Reports of Facilities in Operation Libraries.** University of California, San Diego. Library and Computer Center. Final report, serials computer project. La Jolla, Calif., May 1964, various paging, mimeog.

This report elaborates in detail the development of a mechanized system for handling serials from an original pilot project of seven hundred titles (see *Library Resources and Technical Services*, vol. 7, 1063, p. 71-80) to full operation with three thousand titles in a general library with three major branches. It became necessary to rewrite the original computer (CDC 1604) programmes in the light of improved routines. The value of the report is much enhanced by the inclusion of detailed flow charts and procedures, as well as costing figures (see also *College and Research Libraries*, vol. 24, 1963, p. 489-91).

From the records stored on magnetic tapes six different print-outs are produced periodically, including daily acquisition branch and binding lists, as well as claiming cards. Punching is done from an intermediate serial record on which all incoming issues are recorded. It is very instructive to be able to follow through the very complicated planning necessary to meet the great variety in the features of periodicals so well known to any periodicals librarian.

It is pointed out that it is doubtful whether a library with less than one thousand titles will find mechanization worth while unless computer facilities are available at sub-economic rates. The upper limit for which the reported system would be feasible is around twenty thousand titles. The programmes developed, which can be bought from the University Library, could be converted for use on another large computer and would involve the work of one programmer for about four months.

The general conclusion is that the system is feasible and efficient and that the machine costs are more than compensated for by savings in improved access to information in all departments and units of the library, in bindery routines, and in claiming missing issues. "Experience . . . thus far indicates that even if there was no economic advantage or even if it were necessary to battle for extra funds to maintain the service, it would be worth doing." (JDOC)

10/65–145 **The Modification of an Information Retrieval System by Improving Vocabulary Control, Indexing Consistency and Search Capabilities.** Final summary rept. for 1 Dec. 63-30 Nov. 64, by Edward A. Janning, Mar. 65, 88p. Contract AF33 615 1132, Proj. 7381, Task 738103 ML TR-65-20. AD-613 301. Div. 32 CFSTI Prices: HC$3.00 MF$0.75 Dayton Univ., Ohio. Research Inst. See also AD-428 423. Unclassified report.

This report describes the need for and establishment of vocabulary controls for a coordinate indexing retrieval system involving multidisciplinary terminology. Included in the controls is an improved fragmentation system for the handling of organic compounds. The elimination of role indicators from the indexing procedures is also briefly discussed. The system that is described was established to handle the scientific and technical documents that are on file in the library facility of the Air Force Materials Laboratory. Approximately 16,000 documents have been indexed and the revised thesaurus contains slightly over 10,000 terms. (author)

10/65–146 **Lawrence Radiation Laboratory IBM 1401 computer Produced and Maintained Library Circulation Records,** James H. Kennedy, Livermore, California, the Laboratory, 1964. Various paging. University of California.

Details are given of a loans record system for reports at the Lawrence Radiation Laboratory. The system utilizes many existing IBM 1401 computer programmes. The following information was kept on one punched card: report number, borrower, report classification, microcard, recall data, outside agency data and dates sent. When a report is returned, a punched card is made with its tape number or line record on the report number printout. At the end of each week punched cards for both borrowed and returned reports are used to update the previous circulation file tape arranged by report number. A copy is made of the new tape so that one tape may be sorted by report number with a sub-arrangement by borrower while the other tape is arranged by borrower with a sub-arrangement by report number. An IBM 7094 computer is used for sorting the tapes: an IBM 140 is used for all other steps. Approximate computer times for the present file of 24,500 outstanding reports are: updating master circulation tape—twenty minutes; copying undating master tape—ten minutes; sorting both tapes using IBM 7094 computer—thirty minutes; printing both sorted tapes—sixty minutes. The system is capable of providing overdue notices and statistics. The fact that the system does not give information about the report which would be useful to borrowers for indentification purposes if they received overdue notices is noted as a possible disadvantage. (JDOC)

10/65–147 **Normal Text Techniques.** J. J. Magnino, Jr., Manager IBM Technical Information Retrieval Center. Paper to be presented at the Second Annual National Colloquim on Information Retrieval, University of Pennsylvania, Philadelphia, Pennsylvania, April 23-24, 1965.

The IBM Technical Information Retrieval Center is organized to provide a centralized in-house information retrieval and dissemination system for the scientific and technological community of IBM. Over a period of years, with the seemingly never-ending proliferation of technical data, it was recognized that the services of an information retrieval center were required. There were many information retrieval techniques that accomplished

computer searching and dissemination with commendable results, but the selection of one system required a survey of the state of the art, an evaluation, and a decision involving many levels of management. This decision was not reached easily as many factors were considered. The major ones may be summarized briefly:

1. In a company-wide reporting system, and indeed in the entire scientific and technical community, it was evident that ordinary English sentences provided the most natural and convenient medium for expressing factual data. Text is a medium that is understood, used, and accepted by scientist, engineer, or administrator for reporting purposes.

2. If the textual data could be processed directly into the computer and meaningfully searched, it would be possible to bypass the usual time-consuming and possibly error-inducing classification, coding, keywording, or structuring problems often involved in I.R. systems.

3. The system employed must be flexible enough to be able to handle historical, current, and future data. Again, by education and experience we are trained to accept textual reporting as this universal bridge. Reclassifications, purging lists or rearranging of files for searching should not be required.

4. As the information explosion continues, we must reduce human intervention and data manipulation. It should be possible to go from original source or printing into the computer in a minimum of time. Automatic preparation of machine input at source, or page-readers in the future, again favored the normal text technique.

5. The data base must be capable of being searched efficiently, rapidly, and economically, as well as meaningfully. Information can become a luxury if high costs or extreme delays are involved for the user. Normal Text searching is a reality and has been demonstrated to be economically feasible, as well as capable of accomplishing meaningful information retrieval searches. The IBM Technical Information Retrieval Center (ITIRC) uses this I. R. technique in its day-to-day operation with excellent results. Although we believe this is a major advance in the state of the art, we recognize information retrieval is still an art and we must work to develop an information retrieval science. Our efforts at the Center are aimed in this direction.

10/65–148 **Machine Processing in a Special Library Vol. 1: A Survey of Case Studies.** Louise Schultz, System Development Corporation, 2500 Colorado Ave., Santa Monica, California, SP-1679/001/00, 15 July 1964.

The range of application of EAM and EDP techniques to the technical operations and service mission of the special library is typified by several summaries. A table of principal features of each system studied aids comparison of the applications. Source literature on each system is cited.

10/65–149 **Automated routines in technical services.** Sievers, Patricia T., and Fasana, Paul J., Hanscom Field, Mass., United States Air Force, Cambridge Research Laboratories, Office of Aerospace Research, 1964. AFCRL-64-70.

This paper discusses the approach to automation taken by the AFCRL Research Library, applicable to libraries in general. A brief review of the current market of equipments and techniques is followed by guidelines this library will use in order to effect "total systems automation."

The AFCRL Research Library monograph processing system is discussed as an application of the techniques described. It covers the preliminary search routines using an MP_3 Polaroid camera, the cataloguing routines making use of various machine-generated authority lists, the bibliographic encoding and catalogue-card generation using tape typewriters, and special-purpose data-processing device, the Itek Crossfiler, developed for this library.

Paper given at the 7th Military Librarians Workshop, Silver Spring, Maryland, October 1963. (JDOC)

10/65–150 **Information Storage and Retrieval System: Computer Aspects and Programs. Project Sharp.** Dec. 64, 91p. NAVSHIPS-0900-001-4000, AD-612 674, Div. 32, 30. Bureau of Ships, Washington, D. C. Proj. SF007 01 03, Task 0404. Unclassified report. Available from Superintendent of Documents, GPO, Washington, D. C., 20402, $0.50, as D211.2:SH2/2. Rept. originally prepared at David Taylor Model Basin, Washington, D. C., Rept. no. 1923.

SHARP is a computer oriented information storage and retrieval system developed to resolve some of the problems inherent in the handling, storage and retrieval of scientific and technical literature at the Bureau of Ships Technical Library. The computer is being used to automate to a high degree; (1) bibliographic searches (2) subject matter searches (3) coordinated searches (combinations of both) (4) issuance of library catalog cards and accessions bulletins (5) control of periodicals and journals (6) other aspects under development, such as complete automatic generic computer searching, automatic posting of descriptive terms by the computer in the indexing procedure, and user interest registers. The system is intended to improve the library's capacity to better serve the user and to function as a management and logistics tool for the library. Details of the computer aspects of the indexing scheme, search strategy, thesaurus, computer programs, present research work are reported. Modifications to the system and future plans are indicated. (author)

10/65–151 **Project SHARP Information Storage and Retrieval System: Evaluation of Indexing Procedures and Retrieval Effectiveness,** Department of the Navy, Bureau of Ships, Washington, D. C. 20360, June 1964 (NAVSHIPS 250-210-3). Available from Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. 20402. Catalog No. D211.2:SH2, Price 35 cents.

The results reported in this study were obtained in the course of research supported by the Bureau of Ships, Department of the Navy under Contract NObs 88417, with Herner and Company, Washington, D. C. The report was prepared by Walter F. Johanningsmeier and F. Wilfred Lancaster, and covers the period March 1963 to June 1964.

This is the final report of a study to evaluate the indexing, coding, query formulation, and retrieval effectiveness of the automated information storage and retrieval system of Project SHARP (SHips Analysis and Retrieval Project). The study involved indexing documents according to the coordinate indexing system using links and roles as recommended by the Engineers Joint Council (EJC). The indexing was then coded for keypunching and placed on the computer at David Taylor Model Basin. Sample questions, based on documents in the collection, were obtained from technical personnel at the Bureau of Ships and, from these, queries were coded for submission to the retrieval system. The search results were then analyzed to determine the retrieval effectiveness of the system.

The study indicated that the subject matter of the Bureau of Ships Technical Library can be indexed using the EJC roles, that some of the coding procedures caused a high rate of coding errors, and that the system has high discriminating power, capable of retrieving a small subset of high-relevance documents with very little noise. (authors)

10/65–152 **Description of Cataloging and Indexing System for the ACM Repository.** Morris Rubinoff and John F. White, Jr. January 25, 1965. 63 pp.; Contract AF49 638 1421. AFOSR 65-0424. Unclassified report. Technical rept., AD-612 277, Div. 32, CFSTI. Prices: HC $3.00 MF $0.75. Moore School of Electrical Engineering, Univ. of Pennsylvania, Philadelphia.

The repository library of the Association for Computing Machinery (ACM) was first proposed in 1961, and subsequently the Moore School of Electrical Engineering was designated as the site. The Moore School established an Advisory Board on Information Retrieval, a research team to provide council in relating the repository to current developments. The Moore School research team concentrated on preparation of an organizational and procedural plan for a mechanized information center with 2 principles: (1) a library must make its information as easily accessible to a customer as any store, and (2) information service should be provided at a competitive price. The system aims to store bibliographic elements, four echelons of index terms for enhancing user convenience, and a number of document attributes for expediting user search. The system has a "multiport" entry plan whereby search may be based on any of the many aspects of the contents or the packaging of documents. The concept of the multiport entry is an outgrowth of The Moore School Multilist system (AD–270 573) and has been incorporated into the indexing system via "Added Info" groups. The scheme also permits stochastic search through the system, i.e. attaining the most applicable word from

an initial approximation and a search through related categories, sections and/or classes.

10/65–153 **A Central Information Retrieval System.** D. L. Armstrong and M. T. Grenier. JCHD, v. 5, No. 2, May 1965, pp. 99-102.

Describes a centralized technical information retrieval system for a technically oriented company having over 30,000 employees of whom 6,000 are engineers. Presents Aerojet-General's solution to the problem of the design and establishment of an efficient, economical information retrieval system which would permit the maximum utilization of technical information generated at each of the seven company plants.

10/65–154 **The MEDLARS Project at the National Library of Medicine.** Charles J. Austin. LRTS, 9(1): pp. 94-99. Winter 1965.

10/65–155 **Catalog Subject Searches in the Yale Medical Library.** Brooks, Benedict and Kilgour, Frederick G. CRLI, vol. 25, no. 6, November 1964, pp. 483-7.

An investigation of the use of subject cards in the catalogue of Yale Medical Library was made as an aid in designing a catalogue computerization project. The period of observation chosen was one week during the period of heaviest use of the library during October-November. This period had been determined by plotting the loans records for the previous year. The catalogue is a dictionary, one consisting of a single alphabet, and is in a card form. This is housed in a single cabinet with drawers on two sides. An observer recorded traffic to the catalogue distinguishing between library staff of the Cataloguing and Acquisition Department (the "staff") and the library users plus other library staff ("the public"). The number of trays consulted by a user was also recorded. People were interviewed immediately after they had looked at the catalogue to determine which approach had been usd, i.e. subject, author, etc. In analyzing the data, the unit used was the "search," the number of people involved is not considered. It was found that the peak traffic occurs between 2 and 3 p.m.; the library being open from 8.30 a.m. to midnight. One, two, or three drawers were usually consulted but the record number was thirty. Five hundred and one searches took place during the observation period; these broke down almost equally between the "public"—251 and the "staff"—250. Of the total number of searches, 12.8 per cent were subject ones; 17.9 per cent of the searches by the "public" were subject ones. Results also indicated that the "public" seeks material in English almost exclusively; there is a heavy preference for finding as recent material as possible; a considerable number of the "public" choose to locate a known item by the subject approach; some people consulted a subject in the catalogue to learn the classification of a topic, i.e. use it as a quick guide to the location of material on the shelves; a particular form of material is often sought in subject searches, e.g. periodicals of biochemistry, textbooks of embryology.

J. Condon and C. E. Dent. SPLB 55 (10), pp. 703-705 (Dec. 1964).

To speed processing of book purchase orders the Library of Systems Development Corporation has developed a special purchase order to be handled by the library. A blank check is part of the form and has printed on its face "This check has a limit of $25.00." (MCP)

10/65–157 **Library public services in the age of data processing,** Howard W. Cordell. *Florida Libs.*, 15 (2) June 1964, 11-14.

The new Florida Atlantic University Library is to be organized so that as many processes as possible may be performed by computer. An unfamiliar designation on the library staff is Head of Liaison Services, a post held by the writer, whose responsibilities cover the link between library, teaching, and research, and who will use the computer to help with curriculum planning and information. Other senior staff are appointed to supervise information retrieval and data processing for library records. The computer is used for printing catalogues, for producing information to support individual courses, for document retrieval based on the facts stored during the cataloguing operations, for informing reference librarians of the progress of items being acquired for stock, for listing serials and for detailing information of books on loan. Book issues are to be controlled by the IBM 357 data collection system. This involves an identity badge for each library user, a book-card for each volume with details in print and in machine code, and a machine which registers these two at the moment of issue and produces two cards; a charge card for retention in the library, and a date-due card for the reader. Each day's issue is converted into a printed list from the charge cards. On return, the date-due cards are removed from the books and used by the computer to amend the current list of books on issue. Several tasks, including the sending of overdue notices, can be taken from the library staff and performed, with the aid of the computer, by the university business office. (LSCA)

10/65–158 **Mechanized production of bibliographical lists,** R. E. Coward. *Lib. Assn. Rec.*, *67* (2) February 1965, pp. 50-51.

Briefly describes the IBM Executive typewriter, an electrical, proportionally-spaced machine now being used at *B.N.B.* to produce indexes. The writer recalls the article by Cayless and others (*see LSA* 14858) in which a similar task was described using an IBM 1401 computer; comparisons are made and costs quoted. (LSCA)

10/65–159 **ALPHA, et al,** Croxton, Fred E. *STNW,* vol. 18, no. 3, Fall 1964, pp. 79-82.

ALPHA (Automated Literature Processing, Handling, and Analysis) is a system developed by the Redstone Scientific Information Center, Redstone Arsenal, which serves between six thousand and eight thousand scientists and engineers scattered over an area of more than a hundred square miles. The system is an integrated one covering the administrative aspects of library work and also information retrieval. It consist basically of four tape files operated on IBM 1401(8K), IBM 1410, and IBM 7010 computers without random-access or remote-inquiry devices. The four files are made up of a patron file giving information about the people that the library serves, a periodical file, a book file, and a document file.

Three principles kept in mind were: open-endedness—ability to accept data from other systems as freely as formats and conversion routines permit; the "hopper" approach—permits maintenance and updating without batching by transaction types, use of feedback loops of machine-and man-readable output. There are subsidiary files; e.g. in-process files for the book, periodical and document files, inventory files for the book, and periodical files. The inventory file has an entry for every copy of every title. Disposition, addition, and loan information is posted to the file. The loans record is extracted from this file; it is arranged in two sequences, by call number and by borrower's name. The in-process file is a by-product of the original ordering process. Work is still in progress on the system. (JDOC)

10/65–160 **Computerized Serial Records.** Don S. Culbertson LRTS 9(1): pp. 53–58. LRTS. Winter 1965.

10/65–161 **Automated Circulation Control at Southern Illinois University.** Dejarnett, L. R. *Education Data Processing Newsletter,* April 1964, pp. 7-11.

An IBM 357 data-collection system is used for the loans record at Southern Illinois University. There is a master IBM card for each book with the book identification number punched in it and a plastic identification card for each borrower with a borrower's number punched in it and other particulars about the borrower embossed on it in ordinary writing. When a book is borrowed the date it is due back is keyed-in on the manual keyboard of the IBM 357. The identification card and the book master card are inserted into the input station. This activates the keypunch to which the IBM 357 is coupled and results in the production of two cards, known as Return and Charge cards. Each contains the book identification number, the borrower's number, and the date the book is due back. The Return card and the book Master card are placed in the book pocket and the borrower's identification card is returned to the borrower with the book. The Charge Card is retained at the circulation desk. The Return card is evidence to the checker at the door that the book has been charged out. It is removed at the circulation desk when the book is returned. At the end of each day Charge and Return cards accumulated at the circulation desk are sent to the University Data Processing and Computing Center where they are used to update a tape record of outstanding circulation. This is done using an IBM 1401 computer. Automatic overdue notices are prepared. Other university records which are maintained on

magnetic tape can be used with circulation records to make special studies such as student reading habits by class, course and sex, etc. (JDOC)

10/65–162 **Preparation of a Cumulative List of Periodicals by the Photo-Hollerith Method,** Deszo, L., and Dobay, A., *Racionalizace Prace Stojove Zpracovani Informaci v Knihovnach a Utvarech Tel* (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers), No. 2, 1963, pp. 66–71. In Czech.

The periodicals in question are those acquired by the Library of Hungarian Heavy Machine Industry in Budapest in 1961 and 1962. The organization of work methods necessary for preparing the lists for machine processing are described. Practical examples are given of codes used in individual parts of the listing, prepared with the help of an alpha-numerical IBM tabulator. The collection, mechanically processed and arranged in alphabetical order by titles and countries of publication, is reproduced by a photo offset printing method. (DOC INC)

10/65–163 **Studies Related to Catalog Problems.** Henry J. Dubester. LIBQ, Vol. 34, No. 1, January 19, 1964, pp. 97–105.

This paper is essentially a discussion of the studies of the catalog which were made at the Library of Congress in relation to the automation survey in that library. This survey, supported by a grant of funds from the Council on Library Resources, initiated in the spring of 1961, and presently verging on completion, has as its primary focus the question: Is automation of the information system in a large research library technically feasible? After an initial exposure to the organization and operations of the Library of Congress and a few other large research libraries, the team of experts that pursued the survey focused on two apparently basic questions: (1) Can memory or machine store of sufficient size and with a sufficient rate of response to serve a large research library be developed? (2) Can there be the sophisticated interaction between a human user and the machine memory to permit the type of dialogue which is ever present in the process of using the bibliographic apparatus of the large research library?

In connection with these efforts, certain questions arose regarding characteristics of the dictionary catalog in the Library of Congress; they could be answered only by particular studies since no existing information applied to these questions.

This paper will try to describe the condition of the conjectured system which gave rise to a specific question, then to describe the study designed to satisfy this question, together with the actual results.

10/65–164 **The ASM Information Retrieval System: After Cranfield.** Marjorie R. Hyslop. JDOC, Vol. 21, No. 1, March, 1965. pp. 27–42.

The ASM Information Retrieval System is currently in process of redesign to incorporate major refinements and improvements resulting from five years of operating experience as well as experience and research of others in the field. Principles and methodology underlying these changes are explored for various components of this system, principally vocabulary, links and roles, and computer hardware and programming. The major change is from semantic code to thesaurus as the system vocabulary, and the relative efficiencies and inherent capabilities of each are compared and related to the general principles of vocabulary control.

10/65–165 **Initiating a Mechanized Union Catalog for Medical Libraries in Metropolitan New York.** J. W. Felter, SPLB, 55(9), p. 621, Nov. 1964. A Union Catalog of Medical Periodicals was prepared showing the holdings of 70 libraries. An expert programming consultant was employed to design the system for computer equipment which is available by renting needed time. The system can be kept up to date by regular library personnel. (MCP)

10/65–166 **The Library of Congress Project.** Gilbert W. King, LRTS, 9(1): pp. 90–93. Winter 1965. Summarizes the general requirements for automating the Library of Congress, published in detail as *Automation and the Library of Congress,* 1963. (D.C.W.)

10/65–167 **Putting the Catalog of a Small Company Library into the KWIC Index: A Pilot Study,** A. F. Laubach. SPLB 55(9), pp. 619-620 (Nov. 1964). The KWIC index offers a means of making the library catalog available in different sections of the company. For the first run about 825 cards were processed and printed out as a bibliography listing, a key-word title listing and an author index. Estimated cost of indexing the 60 drawer file is $2000. (M.C.P.)

10/65–168 **Computerized Circulation Work: A Case Study of the 357 Data Collection System.** Ralph E. McCoy, LRTS, 9(1): pp. 59–65. Winter 1965. Describes system in use at Southern Illinois University with six input stations and nighttime 1401 processing of an up-dated circulation list. (D.C.W.)

10/65–169 **The United States Library of Congress Automation Survey,** Barbara Evans Markuson. UNDC, 19(1) January-February 1965, pp. 24-34.

The author's aim is to provide additional insight for those who have read the report of the survey (*see LSA* 14420) and to stimulate the curiosity of those who have not. She includes a list of the survey's main findings and briefly describes reactions to them under the following headings: (i) Justification of proposed expenditures. A mechanized library must justify itself financially, but it is extremely difficult to project the costs and values of a library system which does not yet exist. (ii) Cost of patron time. It is economically desirable to increase the effectiveness of time spent by readers in the library as well as the effectiveness of (iii) staff time. (iv) Centralized processing. The savings in other libraries' cataloguing budgets made possible in 1963 by LC's Centralized Cataloging Service are estimated to be $22m. The survey report shows how this figure might be reduced by increased centralization and automation. (v) Information retrieval. For some time to come the emphasis here will be on the automatic retrieval of bibliographical information. (vi) The large versus small approach. It would not be feasible to test systems in a small library and then scale them up to install them in large ones: in any case the initiative would have to come from the larger libraries. LC's Office of the Information Systems Specialist was established in 1961 to serve as a focal point for the study and development of computer applications to library operations, and methods of producing cataloguing copy in machine-readable form are already being investigated. LC regards the report as a beginning rather than as a conclusion to its interest in automation. (LSCA)

10/65–170 **The Prospects for Mechanization.** Morse, Philip M. CRLI, vol. 25, no. 2, March 1964, pp. 115–119.

Unless large university libraries in the U.S. soon begin to mechanize, their operating effectiveness will deteriorate rapidly. Prospects for mechanizing the following classes of activity are considered: 1, handling of the present contents of the library, e.g. circulation, keeping an inventory of the book stock, physical maintenance of stock; 2, bringing new material into the library, e.g. accessioning, marking in periodicals, cataloguing; 3, assistance to library users, e.g. reference work, maintenance of catalogues and book lists; 4, making material more readily available, e.g. providing microfilms and other copies. As these activities are interconnected, improvements in one may help the others. Mechanization of one class of activity must be planned in relation to the other classes.

There is a sudden change in character of a library when it grows beyond a certain size. At some stage the librarians no longer has control over his material and no longer has contact with the users.

At MIT, science library users do not find material at the place indicated in the catalogue in about one-quarter of their attempts. In this case the "frustration factor" is said to be about one-quarter; it has roughly doubled over the past five years. Not finding material may be the result of the material being in use, at the binders or repairers, mis-shelved, not yet shelved, or lost. A "frustration factor" of one-quarter is larger than it should be. If the "frustration factor" rises above one-half, the library rapidly loses its value as a repository of accessible reference material. To know of a "lost" book before it is required, a large library needs a procedure which will enable the library staff to know accurately and immediately "how the system is doing." The mechanization of class 1 activities is necessary for this. Costs may be reduced by co-ordinating this with mechanization in other departments of the university. Consideration of the mechanization of class 2 could

begin once plans for class 1 mechanization are under way. Data generated by class 1 operations will provide useful and necessary inputs to class 2 operations, e.g. lists of "lost" books can be scanned to decide which should be re-ordered. Many class 2 operations cannot yet be mechanized, e.g. deciding what to order, preparing the original catalogue card. The basic difficulty in the mechanization of classes 3 and 4 is that information printed on paper is more convenient to read and use than the same information projected optically or electronically on some screen. Better, cheaper, and quicker copying processes are required to provide copies on paper of electronic catalogue entries and of microfilmed material. (JDOC)

10/65–171 **Computerized Cataloging: The Computerized Catalog at Florida Atlantic University.** Jean Perreault. LRTS, 9(1): pp. 20–34. Winter 1965.

10/65–172 **Data Processing Aids in Acquisition Work.** Louis A. Schultheiss. LRTS, 9(1): pp. 66–72. Winter 1965. Brief description tion of the acquisitions program of the University of Illinois, Chicago Undergraduate Division, designed for use with a 1401. (D.C.W.)

10/65–173 **Mechanization in documentation: the Ministry of Aviation Technical Information and Library Services,** H. F. Vessey. ASLP, 16(11) November 1964, pp. 341–354. References.

Describes the general problems of mechanizing a Document Centre and the work to date in TIL. The possible uses of punched card sorters, card sorters, flexo-print, and tape-typewriters are discussed. The production of *R & D* abstracts is the responsibility of TIL and is used to disseminate information on the report literature collected by the Ministry. Experiments in which *R & D* abstracts was produced by tape-typewriters are described and future possibilities including the production of indexes are outlined. Work being carried out at present on indexing reports is also discussed. Two flow diagrams show the present system of work flow and a proposed system which is considered necessary if a largely automated system is introduced. The author concludes that computer operation of TIL's documentary activities would be advantageous by releasing staff to increase service to customers. (LSCA)

10/65–174 **Automation in the Reserved Books Room,** Weyhrauch, Ernest H., LIBJ. Vol. 89, No. 11, June 1, 1964, pp. 2294–6.

The mechanization of the loans system in the Reserve Room of Brooklyn College Library is described. Eighty thousand fifty-one-column IBM translation cards were numbered and punched consecutively from 1 to 80,000. Students fill out a standard eighty-column unpunched IBM call card which is taken with the book to the charging counter. A transaction card is placed in the book's pocket and the same number that appeared on the transaction card is stamped on the call card with a Simplex machine. Time or date due is put on the date-due slip opposite the book pocket. The call cards are accumulated in numerical order in bins by the Simplex machines. As books are returned, their transaction cards are pulled from the book pockets and set aside in a box. All the call cards and transaction cards are taken down to the IBM room at a specific time. The transaction cards are sorted into numerical sequence on the 082 sorter. The call cards are punched on the key punch and arranged on the sorter in numerical sequence. Both sets of cards are matched on the collator and those which do not match are checked to see if they are overdues. (JDOC)

10/65–175 **An Alphabetic Index as a Byproduct of Computer Coordinate Indexing.** I. Zarember. *Parameters of Information Science,* pp. 467–470. (see 4/65–1).

In the American Petroleum Institute's computer based coordinate indexing system a printed Alphabetic Subject Index is produced as a by-product. The over-all system uses standard terms for indexing. The indexer analyzes each document and selects the suitable standard terms. From among these terms for the particular document he also chooses several which are primary access terms; these are the ones that are likely to be looped up in an alphabetic printed index. Each entry in the Alphabetic Subject Index contains the alphabetized term, additional amplifying terms, the document title and its serial number of association number. The alphabetic index can be produced almost as quickly as a permuted-title index. It is better than the latter type in that (1) it provides somewhat more information, and (2) the use of a standard vocabulary makes searching relatively rapid and reliable.

10/65–176 **A Specialized Documentation and Retrieval System.** Henry G. Higley, Department of Research and Statistics, American Chiropractic Association, 201 Palatine Drive, Alhambra, California 91801. April 1965.

A study to determine the feasibility of setting up a system for abstracting, classifying, coding, and retrieving information pertaining to all aspects of the human spine and related structures. A relatively small, highly specialized system is needed. The system must be able to retrieve information as well as data. It must be flexible, capable of future expansion, and economical. It would be advantageous if the material could be retrieved on at least two levels: (a) material directly related to the question, and (b) material indirectly or partially related to the question.

The first part of the program consists of determining the function that the information must perform, areas of inquiry, and the frequency with which the various types of inquiry occur. The data are being obtained by reviewing the requirements of the Department of Research and Statistics in the past 3 years and requesting similar information from the other departments and agencies of the National Chiropractic Association.

Secondly, a study is being made to locate the principal sources of information. It is estimated that some 1,600 documents directly related to the subject area are published yearly; about one-half of them are listed in *Index Medicus.*

When the project is completed, the information obtained will be used for the design of a system of classification and coding and the preparation of a thesaurus.

The result of our study indicated that the development of a documentation and retrieval system is a must. A program was established which is functional and not only has served the needs of this department but has been of value to other departments of the A.C.A. as well as to various other groups, institutions, and individuals as judged by the requests we are receiving for information and bibliographies.

Up to August 31, 1964, there were 4,274 abstracts in the collection with a yearly input of approximately 800 documents. For convenience the documents are listed in five groups. (author)

10/65–177 **Automatic Classification for the ASTIA Mathematics Collection.** Barry Zimmerman. May 64, 48pp. Rept. no. 64–22. Technical rept., AD–457 615L Div. 32. Moore School of Electrical Engineering, Univ. of Pennsylvania, Philadelphia. Contract Nonr55140 Unclassified report. Notice: All release of this document is controlled. All certified requesters shall obtain release approval from Office of Naval Research, Washington, D. C. Attn. Code 437.

10/65–178 **MADAM: A User-Oriented Information Processing System for the IBM 1401.** William O. Crossley and Kendall A. Hinman. 8 Jan. 65, 39p. Rept. no. TM2198 000 00. Unclassified report. D—456 940. Div. 30, System Development Corp., Santa Monica, Calif. Technical memorandum.

MADAM (Moderately Advanced DAta Management) is a program which builds and manipulates files, and generates reports. This document describes the system—its rationale, utilization, grammar and vocabulary. Examples of its use in the California Hospital Utilization Research Project (HURP) and at SDC to support the individual scientist in more effective maintenance of his personal document files are given. (authors)

10/65–179 **Report on Applying the SfB system to communications within the building industry.** *RIBA Journal,* vol. 71, no. 9, September 1964, pp. 400–3.

This is the report of a joint delegation of the Royal Institute of British Architects, the Royal Institute of Chartered Surveyors, and the National Federation of Building Trade Employers which visited Bjorn Bindslev in Copenhagen in September 1963 to study and assess the work that he was doing on the application of the SfB-System to contract documentation, work drawing, bills of quantities, specifications, etc. Bjorn Bindslev calls his method Coordinated Building Communication (CBC). He has extended the SfB-System to take advantage of the potentialities of the com-

puter. He uses what he calls 'The Full SfB-Code', this involves always citing the facets in the same order and always taking a symbol from each of the three basic tables.

e.g. (C– fg2 for –: BRICKS: CLAY

whereas in ordinary library classification this would be written fg2.

Where further subdivision is required a sequential number is added after a point representing a fourth facet.

e.g. (1) fg.2.1234

Bjorn Bindslev was joined in practice by his brother Knud Bindslev, a computer specialist. Use was made of the computer facilities at the Datacentralen in Copenhagen. Taking off and coding is done simultaneously from drawings or other documents which may or may not be classified and coded by the architect. The information is punched on to cards which are used to instruct the computers. An IBM 1401 computer may be used to read, punch, and print out. A faster and cheaper way considering the rates charged by the Datacentralen is to use an IBM 7070 computer for printing out. The cheapest way of storing information is said to be on magnetic tape: the tapes costing nothing if left in the Datacentralen. The coding method developed provides for the possibility of some 6,250,000,000 descriptions of items. Bjorn recognizes that at present the system may be able to produce more information than architects and builders know how to use, but he sees this as a problem of education. It is concluded that Bindslev's methods and the procedures that he is evolving might provide a means of co-ordinating and communication between all members of the team in the presentation of contract documentation and in management. Recommendations by the delegation to he Joint Committee on Technical Information of the RIBA, RICS, and NFBTE are included in this report.

A series of articles has appeared in the *Architects Journal,* during 1964 giving details of Bindslev's work at a later date. (JDOC)

10/65–180 **A Program System for the Automatic Issue and Revision of Illustrated Parts Catalogs.** W. H. Maurer and M. L. Reynolds, Feb. 1964. IEEI vol. IECI–11 No. 1, pp. 71–78 (also in IEAI 83 (72): pp. 137–142, May 1964).

The illustrated parts catalog issue-revision system is a group of computer (IBM 7090) programs that can be used to generate a new parts catalog or revise an old one. The basic functions of the system are to process all input data, to merge input data with a history tape of the previous level catalog in case of a revision, and to paginate the catalog. Pagination procedure includes placement of illustration pages and titles in proper relation to their associated lists; formation of revision pages to accommodate overflow data; placement of list captions; deletion of obsolete lists and associated illustrations; placement of running heads and feet, including full sheet deletion notices and revision date; and the assignment of page numbers. (authors)

10/65–181 **Computations Aid Oligonucleotide Analyses.** *CENE 42,* pp. 26–27 (December 28, 1964). Special computational methods are being put to work by a group of scientists at the National Cancer Institute, National Institutes of Health to determine composition and purity of oligonucleotides. Computer techniques speed analyses of oligonucleotide mixtures up to the tetranucleotide level. (H.L.W./A.B.J.)

Reports of Proposed Facilities—Libraries

10/65–182 **The Recording of Library of Congress Bibliographical Data in Machine Form.** Lawrence F. Buckland. A report for The Council on Library Resources, Inc. Inforonics, Inc. P. O. Box 207, Maynard, Massachusetts, November 23, 1964, Revised, February 1965 Washington Council on Library Resources, Inc., 1965, L.C. 65–18647

This report describes the results of a study of a practical method of preparing Library of Congress card catalog data in machine form for (1) the automatic typesetting of cards and book catalogs, and (2) distribution to other libraries throughout the country for all foreseeable bibliographic and typographic applications of such data, including preparation of local catalogs.

The first step of the method proposed is to type the card data on a perforated tape typewriter in a way which identifies all of the items on the card. After the data is edited and corrected, it is processed by a computer to form (1) catalog card typesetting tapes, (2) the National Union Catalog, (3) Library of Congress books, Subjects Catalog, magnetic tape catalog files, and a master magnetic or perforated paper tape record copy for distribution. The procedure also allows the recording of data which does not now appear on Library of Congress cards, should studies find that the value of the data exceeds the cost of recording it.

A demonstration was performed in which cards were typed to produce perforated tape records. These record tapes were automatically converted to a variety of output forms ranging from phototypeset catalog cards and book catalog entries to tape typewriter and line printer produced catalog cards.

10/65–183 **Reference Data for Government Contractors.** R. S. Huleatt, *SPLB 55* (10), pp. 680–687 (Dec. 1964) Steps are given for setting up a central control office for the government reports and specifications required for government contract bids and operations. A list of reference sources is included and suggestions made on maintaining the collection. (MCP)

10/65–184 **The Library of Congress Project.** Gilbert W. King LRTS 9(1): pp. 90-93. Winter 1965. Summarizes the general requirements for automating the Library of Congress, published in detail as *Automation and the Library of Congress,* 1963. (D.C.W.)

10/65–185 **Automatic Classification Method for Publication, Library and Documentation Centers**—Von Martin Scheele. NATW, v. 52, no. 1, Jan. 1965, pp. 1–10.—Automation of documentation is best applied to the first bibliographic stage, that is to the treatment of the title with the help of punched tapes in combination with electronic calculators. The outstanding feature of the method described is the automatic classification prepared by a suitable classification system where—for instance in biology—the objects are treated separately according to a hierarchical system, and the properties are considered from the points of view of heirarchy and free basic concepts. (In German.) Ein Verfahren zur Automatischen Klassifizierung fur Veroffentlichungswesen, Bibliothekswesen und Dokumentation. (BATR)

10/65–186 **Dissemination of Information.** I. A. Warheit LRTS, 9(1): pp. 73–80. Winter 1965. Briefly describes the use of and techniques for creating common computer-prepared dissemination tools. (D.C.W.)

10/65–187 **Publication Patterns of Scientific Serials.** David Bishop, Arnold L. Milner, and Fred W. Roper. Biomedical Library, Center for the Health Sciences University of California, Los Angeles. AMDO Vol. 16, No. 2—April 1965 pp. 113–121.

American libraries are undertaking computer-controlled inventories of serials holdings. Such automated records are posited on predictability of serials receipts. Publication patterns of serials, falling into 24 distinct classes and two further categories, are developed for machine-based prediction and record updating. The relationship between publication patterns and frequency/dating systems is examined. The publication patterns listed are considered universally applicable to machine methods. The system under development at UCLA uses publication patterns to project the next expected issue of a serial, as well as to prepare forms for bindery and to update both bound and unbound holdings inventories. Although other systems may ignore the problem of bindery, the problem of predictability of future issues is still central, especially when more than one future issue per title is forecast for receipt as is current practice in those systems known to be operational. However named or defined, publication patterns of serials are an essential basis for computer control of serials records.

Reports of Proposed Facilities—Other Information Centers

10/65–188 **The Documentation Service of the Institute of Experimental Medicine and Surgery of the University of Montreal** (Le Service de Documentation de L'Institut de Medecine et Chirurgie Experimentales), G. Ember and G. Gabbiani. 1963, 4p. AD–436 517, Div. 32, 16, Montreal Univ. (Quebec). Grants AF AFOSR62 356; 29 63, AFOSR 64 0683, Unclassified report. Reprint from (Bull. A. I. D.) Int I-II, pp. 19–22, 1963. (Copies supplied by DDC)

The system (SSS) of evaluating, storing and retrieving medical information at the Institute of Experimental Medicine and Surgery of the University of Montreal is described. The importance of the human element in this type of work is stressed and the possibility of mechanization discussed. (author)

10/65–189 **Civil Defense Information Systems Analysis** (A Feasibility Study of Research Information Exchange). AD–612 299, Div. 32, 18, CFSTI Prices: HC$4.00 MF$1.00 Research Triangle Inst., Durham, N. C. Final rept., vol. 1, 1 Dec. 63-15 Jan. 65, by W. T. Herzog and J. E. Jenkins. Jan. 65, 146 pp. Rept. no. R-OU-158-1 Contract OCD PS64 56, Task 4631B, Unclassified report.

A feasibility study was performed of a civil defense scientific and technical information system that will insure the ready availability of information to all pertinent OCD elements and to contract research personnel. A brief analysis of scientific information systems is presented as background for a discussion of the present civil defense information system. The potential sources of civil defense information are listed. The civil defense information system is defined to be the relationships or exchange between the sources of civil defense information and the users of this information. Methods for improving this exchange are presented. These suggestions include the establishment of an information analysis center, the printing of a quarterly technical progress review, and the use of standard report format and indexing procedures. To foster standard indexing procedures, a "Thesaurus of Civil Defense Descriptors" is provided (AD-611 438). (author)

10/65–190 **User-Oriented Information Systems for State and Local Government.** Herbert H. Isaacs. AD-612 997, Div. 32, 26, 30, CFSTI Prices: HC$1.00 MF$0.50 System Development Corp., Santa Monica, Calif. Professional paper, 5 Mar. 65, 23 pp. Rept. no. SP-1988, Unclassified report.

Potential state and local government users of information systems are classified in terms of categories of activities and levels of organizational function. The general information requirements of these users are described and categorized in terms of basic system capabilities. The existing situation in state and local government information systems is then described, indicating a need for more advanced, user-oriented techniques. Such methods and techniques currently available or under research and development are then surveyed. Brief descriptions are given of basic ideas or operating principles, along with reference to more detailed documentation. The paper concludes with several recommendations on how best to maximize transfer of knowledge, evaluate new procurements, and carry out new system design. (author)

10/65–191 **The Qualitative Development Requirements Information (QDRI) Registered Organization Data Bank (RODATA),** James G. Peirce and Charles Beugless. AD-612 595, Div. 32, 18 CFSTI Prices: HC$1.00 MF$0.50 Frankford Arsenal, Philadelphia, Pa., Research and Development Directorate. Mar. 65, 20p Memo rept. no. M65-13-1. Unclassified report.

The QDRI Program is the Army's research and development information-to-industry program, which has been assigned to the U. S. Army Materiel Command for management. QDRI is defined as information (written or verbal) regarding current and future Army requirements for development of materiel exclusive of advanced procurement information. It has become apparent to most of the managers of the QDRI program that there are enough items of data being collected about qualified organizations and from the information interfaces between the Army and Industry to require the development of an automated data system to acquire, store and process commonly needed information in a uniform manner, and in addition certain specialized information for individual Army installations, which data might be useful to other installations under varying circumstances. There is enough evidence to indicate that a QDRI oriented data bank can be designed and be put into partial operation by the end of FY 1965, using existing technology and equipment available to the AMC installations involved in the QDRI program. In the development of the QDRI Managers Guide, the concept of a data bank for QDRI continuously crystallized into the description now presented. This data bank will be an information science oriented operation, using primarily existing operational IBM equipment and the existing AUTODIN/AUTOVON network for rapid transmission of data from each RDTE and procurement qualification activity/installation, and vice-versa.

10/65–192 **Immuno-Chemical and Radiobiological Studies on Mammalian Homotransplantation I. Bibliographical Research.** Annual progress Robert F. Sloan rept. AD-457 063, Div. 16, 32. Army Reserve, Santa Monica, Calif. Research and Development Unit. 1 Jan.-31, Dec. 64, 20 Jan. 65, 9p. Unclassified report.

At the present time many research library personnel are working to synthesize program systems to document, control, and retrieve specialized problems in the medical literature. The Index Medicus does a very good job covering the general medical literature, but it is beyond the scope of Index Medicus and MEDLARS project to cover the medical specialities to the depth needed by research workers in fields such as transplantation. Therefore, it is the specific aim of this project to synthesize a practical program which will permit investigators in the transplantation field to retrieve any given publication references needed. (author)

10/65–193 **Dynamic Simulation for Market Planning,** George B. Hegeman *CENE 43*, pp. 64-71 (January 4, 1965). To provide effective assistance to management, market research and long range planning staffs must combine skill in information retrieval, data analysis and forecasting with sound business judgment. Simulation techniques are particularly appropriate to the problems of chemical market research and planning. Simulated, computer models of industrial marketing systems can help companies set sales goals for present operations or investment plans for new efforts. (HWL/ABJ)

10/65–194 **A Mathematical Model for Mechanical Part Description,** J. Y. S. Luh and R. J. Krolak, *CACM 8*, No. 2, pp. 124-129 (February, 1965).

The flexibility of a mathematical model takes advantage of the common information requirements of computer-aided engineering drawing, numerical control tape generation, and physical characteristic computation. By judicious control of man-machine communication requirements, improved results over conventional engineering design processes are possible. An English-like input language, tailored for use by draftsmen and designers, will describe the part and specify the output desired. One approach to the mathematical model consists of a group of surface-defining quadric equations, which are created by a system of modular subprograms. Other subprograms will convert the mathematical model into instructions for driving automatic drafting machines and numerical controlled machine tools. Physical part characteristics, such as center of gravity, can be computed by subprograms and used in dynamic analysis work. The proposed over-all system is presented and experiments and demonstrations are discussed.

10/65–195 **Mechanizing a Large Register of First Order Patient Data.**—Frederick J. Moore, *Methods of Information in Medicine*, v. 4, no. 1, Mar. 1965, pp. 1-10.—Results of studying the patient register of the Los Angeles County General Hospital indicate that it is feasible to mechanize this file of approximately 1,000,000 cases. It is therefore realistic to contemplate development of interagency health information systems in metropolitan regions. (BATR)

10/65–196 **Formen und Vorgange der arztlichen Diagnosenbildung** (Methods and examples of medical diagnosis) Gerhard Oberhoffer. *NADO, 15* (4) December 1964, pp. 168-173. Diagrs.

Although medical diagnosis is an art, it is an art founded on logical, mathematical and statistical data. In hospital treatment, diagnosis is already based on mathematical data such as pulse rate, blood count, etc. Such data lends itself ideally to mechanical documentation methods. The raw material of diagnosis can be easily adapted to thesauri treatment, symbol programming, and machine programming. The use of mechanical documentation would speed up diagnostic methods and lead to uniformity of diagnosis. Two diagrams are given to show the flow of information from observation of the patient to the ultimate computer interpretation. (LSCA)

Training in Documentation

10/65–197 **Library Education: The Role of Classification, Indexing, and Subject Analysis.** D. J. Foskett. LIBQ, 34 (4) October 1964, pp. 362–373.

The paper is based on the assumption that the study of classification and subject analysis represents the zenith of education for librarianship, though not as an end in itself, but as a means to reader service. Before the twentieth century, scholars knew what literature they required and scholar-librarians would be familiar with the same background. The subject classified catalogue was one answer to the new need for information: as was the Dewey Decimal Classification. Such tools are not to be confused with filing and sorting devices. Definition and correlation of terms are imperative. Reference is made to the surveys at Cranfield College of Aeronautics and Western Reserve University: and to the work of Jean Piaget and his colleagues at the University of Geneva, notably *The early growth of logic in the child*. The traditional classifications are incapable of accommodating concepts unforeseen by their makers. The nature of language is considered in this context, with particular reference to Louis Hjelmslev, Copenhagen Linguistic Circle. Ranganathan, followed by B. C. Vickery, and the Classification Research Group in London, was the first to see that classes should be regarded as sets of elements capable of combination with each other, having regard to the categories of Personality, Matter, Energy, Space and Time. Relationship symbols are mentioned: position in a classification scheme; the "interfixes" of the U.S. Patent Office; the "infixes" of Western Reserve; J. E. Farradane's theories. New mechanical developments include the technique for updating and printing out lists of references e.g. the *B.N.B.* the KWIC indexes; automatic permutation techniques, as in the examples given in *The London Education Classification*. For study in librarianship, comparative analysis of current trends is more fruitful than study of individual schemes in isolation. The new Classification Society in London will provide a forum for discussion. *Comment* by Harold Borko is appended on one of his experiments at the System Development Corporation. (LSCA)

10/65–198 **The Development of a Methodology for Systems Design and Its Role in Library Education.** Robert M. Hayes. LIBQ, 34 (4) October 1964, pp. 339-351.

Information science may be regarded as presenting an addition to library science, not a substitute for it. It is a theoretical discipline with broad design and with system design as its core concept. Since the concern of system design is the analysis of complex situations and the application of theoretical models to them, education for it must present the methodologies involved. The role and nature of methodologies is discussed: the areas in which methods are being developed or needed are: (i) use study; (ii) vocabulary development; (iii) the technical details of internal system processes, especially data processing; (iv) file organization; (v) intellectual problems in judgement of relevancy and screening of material; (vi) performance and cost measurement; (vii) organization relationship. The second and fourth are treated in greater detail. In library education, methodologies affect the understanding of subject cataloguing, classification and filing procedures. Four main areas of study emerge: usage, operation, organization and design. (LSCA)

10/65–199 **The Librarian's Continuing Education.** Carl W. Hintz. PNLA Q., 29 (2) January 1965, pp. 120–122.

Continuing education becomes more important as we are faced with the growing obsolescene of knowledge. The new sciences of automation, documentation and information retrieval are advancing rapidly, yet they are rarely included in degree work. There is a lack of formal organization in the library profession, there is no general certification and membership in professional bodies is voluntary. Courses are needed in scientific management and public administration and in-service training should be on a more formal basis. Many opportunities for continuing education are already available through library schools, associations and conferences and we have a good body of professional literature. We need to establish a pattern of continuing education with both employer and employee realizing the need to make available or have the proper opportunity for participation. (LSCA)

10/65–200 **The Age of the Symbol—A Philosophy of Library Education.** Abraham Kaplan. LIBQ, 34 (4) October 1964, pp. 295–304.

Critical problems affect librarianship in contemporary society, problems of definition of function and of the excess of demand over supply, accentuated by the volume of information and theories about information, together with the improvements in the technology by which information is produced, processed and transmitted. In philosophy, aesthetics, linguistics, psychoanalysis and political science, this is the age of the symbol. The library is a repository for all learning and experience, good and bad, serving society as memory serves the individual; it is a means of education; and an instrument of research, in its genuine sense. The precise function of the library must be defined in relation to the responsibilities of the genuine researcher. Professional education, requiring a period of probably three to seven years, should include: (i) a humanistic basis implying such studies as the sociology of knowledge and the history of ideas in broad cultural terms, incorporating the love of learning, of ideas, of truth and of books implicit in such a vocation, and sense of professional responsibility; (ii) vocational elements, including not only skills, but judgement; (iii) intellectual disciplines, such as logic, linguistics, mathematics and theories of information. Automation must be viewed with balance and control. (LSCA)

10/65–201 **Theoretical Principles of Information Organization in Librarianship.** Vladimir Slamecka and Mortimer Taube. LIQB, 34 (4) October 1964, pp. 352–361.

The development of the "service" concept is examined from the custodial function to the provision of any required items through inter-loan or photocopy. Librarianship mediates knowledge rather than generates it, therefore new techniques and the application of techniques from other fields have made direct impact. Librarianship comprises intellectual content, or principles, and skills: principles formulated through an inductive process are subject to change and further synthesis. Skills and service must react to pressures from within and without. Information, the proper object of library organization, comprises all the elements of the study, in which philosophy, mathematics, linguistics and engineering are also interested. The "topological" and "intellectual," or "physical" and "conceptual" relations of class order are examined in regard to information organization. Criteria must be chosen so that the apply to the entire family of classes, e.g. in the case of information, all other material, as well as books, must be included. For certain types of materials, subject analysis by machine has already obviated subject cataloguing. But the continuing formalization of class relations and criteria, and of verification and improvement remain the human contributions. An outline is suggested for a course in organization of information in librarianship. (LSCA)

10/65–202 **Continuing Education for Librarianship.** W. O. Youngs. PNLA Q., 29 (2) January 1965, pp. 122–126.

Library school curricula must begin to include courses in documentation, administration, cost accounting, personnel administration, budgeting and data processing in order to give the good young professionals emerging into the field some basic knowledge of these important areas. Subject specialization is also important and might be integrated into library science programmes on the German model, in which a three year library science degree follows the Ph.D. Associations, institutes and workshops are invaluable but there is also a great need for more in-service training. Seattle P.L. has established a professional library to encourage staff reading. Denver P.L. has had training sessions, surveying subject departments. But more must be done to encourage the professional to make contact and to discuss his profession and its problems. (LSCA)

10/65–203 **Surveying Russian Technical Publications: A Brief Course.** J. G. Tolpin. SCIE 146, pp. 1143–1144 (1964). Describes a course in which scientists learn to identify the subject matter of Russian publication in eight to sixteen lessons. (K.G.)

Use and User Studies

10/65–204 **Catalog Subject Searches in the Yale Medical Library,** Benedict Brooks and Frederick G. Kilgour. CRLI, 25 (6) November 1964, pp. 483–487. Table, diagr., references.

A study was made of the use of subject cards in the catalogue of Yale Medical Library. The data will be used in the design of a catalogue computerization project at the medical libraries of Columbia, Harvard and Yale. Consultations were

observed during all hours of opening in one week of heavy use. Peak use was between 1400 and 1500 hours. A questionnaire, devised to elicit data on approach by subject or author, revealed that, out of 501 searches by public and staff, 64 were for subjects. The public uses the subject approach more than the staff. The largest number of subjects consulted in a single search was four. One conclusion drawn from the study is that the traditional card catalogue, with an average of about 1.6 subject entries per book, does not furnish references adequate to meet new demands. (LSCA)

10/65–205 **The Study of the Use and Users of Recorded Knowledge,** Philip H. Ennis. *LIBQ,* 34 (4) October 1964, pp. 305–314.

Flexibility of all library functions is imperative, so as to embrace future as well as present users of information, general and specialized audiences. Studies inside the library deal with problems of contact between users and a library function: studies outside the library recognize the library as one among many sources of information. Generalized and specialized audiences may be distinguished by: (i) their motivation to read: (ii) the qualitative and quantitative relation between writers and readers. Certain problems have a direct background of social science, such as the preservation of reading as an adult habit and the diffusion of information though special interest groups. Studies relating to specialized audiences can use the touchstone of efficiency, not appropriate to non-specialist readers. Many user studies have been limited in scope; study of use and users should be related to intellectual roots; specialized groups themselves inter-relate. Possible approaches to the problem are examined: through the training of students in sociology, psychology and economics; by an increase in the numbers of specialists on library school faculties; by higher entrance requirements for students. A table shows the distribution of use studies among library science Ph.D. dissertations 1925-63 compared with use studies carried on in other settings, compiled by R. A. Davis and C. A. Bailey. (LSCA)

10/65–206 **Auskunft und Beratung in der Offentlichen Bucherei** (Information and advice in the public library) Werner Jahrmann. *Eucherei und Bildung,* 16 (2) February 1964, pp. 61–66.

A Swedish survey which examined 2,442 questions asked in nine libraries during six working days is reviewed. 47.7% of these questions could be answered quickly with information or bibliographical data; 9.2% required research on the part of the librarian; 43.1% were answered with advice to the reader. It is questionable whether the conclusions of this survey could be directly applied to German libraries, but no similar survey material is available for Germany. Bare statistics of total reference enquiries are inadequate, but they suggest that the German librarian can devote less time to reference enquiries than his Swedish colleagues. A German survey might be considered in selected large public libraries to examine the nature of enquiries and enquirers, the significance of the decline in readership after the age of 26, and the demands placed upon public libraries to cope not only with bibliographical, literary and technical information, but also with advice on civic and personal matters. (LSCA)

10/65–207 **Documentation and Research Workers as Its Users.** Elin Tornudd. *NADO* v. 15, no. 4, Dec. 1965, pp. 182–186. Surveys studies regarding information needs and uses carried out in the United States, the United Kingdom, and in Scandinavia, and advocates similar studies in German. (In German.) Dokumentation und Forscher als ihre Benutzer. (BATR)

10/65–208 **A Quantitative Measure of User Circulation Requirements and Its Possible Effect on Stack Thinning and Multiple Copy Determination,** Richard W. Trueswell. AMDO, v. 16, no. 1, pp. 20–25, 1965.

The strategy outlined in this article describes a possible aid to the librarian for thinning a library's stacks according to the criterion of user needs. The method uses the last circulation date as a parameter of user circulation requirements. The resulting stack collection would, by design, satisfy over 99 percent of the user circulation requirements and yet be of minimum size. Preliminary, but not yet fully validated research, indicates that the number of volumes in a library's holdings may be reduced by 60 to 70 percent yet satisfy well over 99 percent of the user requirements. It may also be possible to determine which books should have multiple copies in order that user disappointment can be minimized. The effective result of this approach is that there now may be a quantitative method of maintaining the library's holdings at a reasonable level and, in addition, a reduction in the disappointment of the user who is unable to find the book he wants because of use by another. (author)

10/65–209 **Physics Abstracting—Use and Users.** D. J. Urquhart, June 1965. *JDOC* Vol. 21, No. 2, pp. 113–121.

In May 1963 a survey was made of the sources of references to publications requested from the National Lending Library for Science and Technology. This survey, in conjunction with the one carried out by the Advisory Council on Scientific Policy on the information of needs of physicists and chemists, throws some light on the needs of physicists. There is some evidence to suggest that physicists need a current title publication and improved indexes. (author).

10/65–210 **Survey of Information Needs of Physicists and Chemists; the Report of a Survey Undertaken in 1963-4 on Behalf of the Advisory Council on Scientific Policy.** June 1965. *JDOC* Vol. 21, No. 2, pp. 83–112.

The survey assumed that the present conventional information tools and systems will be needed for years to come. Moderate improvement of present services was the goal. The questionnaire (here appended) was distributed to a balanced sample of 6,194 physicists and chemists. Half of these were returned and usable—3% of those active in the United Kingdom. The results are presented in twelve tables with some interpretative comments on the data. A notable finding was the similarity in the information seeking and using habits of the two groups. Contrary to earlier surveys, there was considerable use of the literature and there was no general satisfaction with the methods of getting information. Personal contacts are not overwhelmingly preferred as had earlier been reported. Abstracts and original papers are the most important sources; patents and reports are the least frequently used. The use of review articles is almost universal. The library/information department is fairly well regarded but as a passive stone rather than as a dynamic center of information. A need was expressed to grade papers to indicate what is new, what is introductory, and what is experimental. Five suggestions are given for future investigations, including the comment that large-scale trials of new publications, particularly those presenting alternative forms of original papers, should be seriously considered. (D.C.W.)

Equipment and Material Descriptions

10/65–211 **Automatic Search Systems.** A. A. Fel'dbaum. (Entsiklopedia Izmerenii,) Kontrolia i Avtomatizatsii No. 3: pp. 57–62, August 1964.

The paper presents a brief dicussion of the fundamentals of automatic search theory. The first section is devoted to a definition and description of self-adaptive and self-teaching automatic search systems. Automatic optimalization systems are discussed as an important subclass of automatic search systems. "Blind" search and search with an analysis of intermediate results are discussed. The second section analyzes the various types of automatic search systems. The method of steepest descent and the gradient method are briefly discussed; electronic automatic optimalizers and their application are discussed from the point of view of noise problems and current computer techniques.

10/65–212 **The Hardware of Data Processing.** C. D. Gull. *LRTS,* 9 (1): pp. 6–19. Winter 1965.

10/65–213 **Basic Information on Business Machines.** No Author, *Podnikova Organizace* (Factory Organization), No. 10, 1964, pp. 468–469, Continued from No. 9. In Czech.

Methods and hardware used in the field of data-recording and data-processing are discussed. Emphasis is placed on the keypunch and other machines manufactured by Aritma Enterprises. The layout of the two-part, 90-column cards and the punching code used by Aritma are described. The processes of punching, sorting, reproducing, verifying, and tabulating on the Aritma 220

electronic sorter with an operating speed of 60,000 cards per hour, the Aritma 140 numerical keypunch machine, the Aritma 400 keypunch machine, and the Aritma 200 sorter with an operating speed of 24,000 cards per hour are described. (T.G.F.)

10/65–214 **The Microimage Characteristics of a Kalvar Film,** M. E. Rabedeau, *PSEN* 9, No. 1, pp. 58–62 (January/February, 1965).

The microimage characteristics of Kalvar film, which differ from those of silver halide materials, are given. A "viewing adjacency" effect has been observed, and granularity has been found to be a nonmonotonic function of density. The modulation transfer function (MTF) of the film has been found to be very nearly independent of exposure and viewing adjacency effects: consequently, image quality can be specified by a unique MTF. MTF data are presented for the particular Kalvar film investigated.

10/65–215 **Microfiche Is "In" for Good Reasons.** *SSTM*, v. 6, No. 4, May 1965, pp. 45 + 4 pages. Discusses advantages of microfiche for document storage. (BATR)

Special Processing Languages

10/65–216 **NPL: Highlights of a New Programming Language,** G. Radin and H. P. Rogoway, *CACM* 8, No. 1, pp. 9–17 (January, 1965).

NPL is a new, high-level program language designed to allow both the novice and the experienced programmer to program in a straightforward manner a broad range of applications for contemporary and perhaps future computers. This paper discusses the design criteria for the language and is an examination and justification of those features which are not generally present in higher-level languages and of other features for which the NPL approach is somewhat different from that in previous languages. Topics discussed include program structure, types of data, input/output, non-sequential facilities, and compile time facilities.

Coding and Notation

10/65–217 **An Evaluation of Links and Roles as Retrieval Tools.** Stanley M. Cohen, Carol M. Lauer, and Bettina C. Schwartz. *JCHD*, v. 5, No. 2, May 1965, pp. 118-121. Describes the information indexing and retrieval system used at Linde Division, Union Carbide Corporation. (BATR)

10/65–218 **A Cluster of Algorithms Relating the Nomenclature of Organic Compounds to their Structure Matrices and Ciphers.** G. M. Dyson. *ISSR*, vol. 2, No. 3, December 1964, pp. 159-99.

General methods have ben devised for the computer conversion of the names of organic compounds to structural matrices from which the notation and molecular formulas can be automatically deduced. The procedures involve the systematic use by a cluster of algorithms, of the stored data concerning the nonsystematic morphemes and semantemes of organic nomenclature. (JDOC)

10/65–219 **The "Basic Structure — Substituent — Connector" Concept for Handling Organic Compounds for Information Retrieval Purposes.** Panos Kokoropoulos and Frederick Scheffler. *Parameters of Information Science*. pp. 399-402. (see 4/65–1.)

A system is proposed for handling organic compounds in a manual/computer coordinate indexing system, multi-disciplinary in nature, with a potential of 100,000 documents.

The proposed system is based on the concept "basic structure—substituent—connector," which makes possible the handling of a large number of organic compounds with a relatively high degree of specificity, while maintaining a "controlled" vocabulary.

The vocabulary is based on accepted I.U.P.A.C. nomenclature. The identity of basic structures and characteristic organic groups is maintained.

10/65–220 **The Generation of a Unique Machine Description for Chemical Structures**—A Technique Developed at Chemical Abstracts Service.—H. L. Morgan. *JCHD*, v. 5, No. 2, May 1965, pp. 107-113. (BATR)

Specific Classification and Indexing Systems

10/65–221 **The Use of Punched Cards for the Production of Multiple Copies of an Alphabetical Subject Index to the UDC.** C. F. Cayless and F. H. Ayres. Aldermaston, the authors, AWRE Library, 1964. 7p.

Work done to produce multiple copies of the alphabetical subject index in the AWRE Library using a punched-card system is described. The first attempts made on an IBM 421 accounting machine were soon abandoned. The work was continued on an IBM 1401 computer. Two punched cards were used for each entry, one for words and one for UDC number. This was primarily because the entries, compiled by chain procedure, frequently contained more information than could be put on a single eighty-column card. This meant that the entries had to be numbered serially. This, however, meant that the cards could be sorted more quickly as numerical sorting is quicker than alphabetical. The four thousand entries were numbered using a six-digit numbering sequence so as to allow for the insertion of new entries, e.g. entry 024800 was followed by 024900. The punched cards themselves had to be numbered because of "see also" entries on the UDC side. The entry was printed in ordinary language along the leading edge of each punched card. Each batch of cards was checked in the library after punching for coding and punching errors against the prints-out for the batch. Machine verification was unnecessary. A complete updating print-out was made every quarter, or after 250 additions and amendements. The use of the printed subject index in other UKAEA libraries is discussed. Eighteen months of use has shown that the original aim of saving senior staff time by having multiple copies of the alphabetical subject index has been achieved. Also the book-form is quicker and simpler to consult than the card index—tests showed that it took up to three seconds less to consult an entry in the book-form index than it did to consult one on a card-wheel (itself quicker to use than a conventional card catalogue.) Future plans include the use of magnetic tape in place of punched cards. This will have the advantage of doing away with the need for two cards per entry; allowing much more information per entry; being able to print out at least twice as fast as at present; being able to make an amendment without the manual handling of a card; making permuted indexing of the alphabetic heading possible; and making a sort in order of UDC number possible—this would enable classifiers to see what alphabetical headings had been used for a particular number. The cost of producing five copies of a five-thousand-entry index, assuming that a print-out is made after every five hunderd additions, i.e. ten prints-out in all, was estimated to be £25 10*s*; the cost of these indexes by typing was estimated to be £320. The annual cost of producing forty copies of the subject index by computer, assuming eight prints-out and that the total number of additions and alterations is two thousand per annum, was estimated to be £495 assuming labour costs 25*s* per hour and £382 assuming labour costs 15*s* per hour.

An earlier edition of this article prepared during 1962-3 appeared in the *Library Association Record* in October 1964. (JDOC)

10/65–222 **The Progress of the Universal Decimal Classification in the USSR.** A. A. Fomin. REDO, vol. 32, No. 2, May 1965, pp. 54-55.

Description of progress made in the USSR in introducing the UDC since 1963, when its use became obligatory for scientific and technical publishers of books and periodicals, information centres and special libraries. Publication of UDC editions. Activity of the Interdepartmental Commission for Classification. (REDO)

10/65–223 **Computers and Classification Systems.** Robert R. Freeman. JDOC 20(3): pp. 137-45, Sept. 1964.

Published experiences dealing with the use of data processing equipment in conjunction with classification systems, particularly the UDC, are reviewed. Earlier conclusions, generalized from experience with punched card equipment, were too pessi-

mistic. Difficulties in handing UDC notation in computer systems and methods for overcoming them are discussed. Ways in which computers can aid in studying, improving, and using classification systems are discussed. Experience at Meteorological and Geoastrophysical Abstracts is cited. (author)

10/65–224 **SYNTOL. Vol. II.** J. C. Gardin. Rutgers Series on Systems for the Intellectual Organization of Information. S. Artandi, ed. Rutgers University Press, 1965. Paper $3.50.

A description of SYNTOL is given according to the pattern which was developed for the Rutgers seminar series on the "intellecutal organization of information." SYNTOL was designed by J. C. Gardin, in France, as a general system for information processing which relies on human indexing to provide the input and on a man/machine combination to produce the required output. The system is general in the sense that it is not limited to a specific field of application and it is flexible because it can be applied at different levels using different storage and retrieval equipment. The intellectual content of materials is expressed in terms of the index language of SYNTOL which is created empirically. The indexing terms belong to two kinds of organization, the paradigmatic or vertical organization and the syntagmatic or horizontal organization. When searching, natural language inquiries are re-phrased in the SYNTOL language and matched against the stored document representations. The system can be used both in a pre-coordinate and a post-coordinate way. (S.A.)

10/65–225 **Some Questions Regarding the Classification of Metallic Materials,** Kofnovec, L., *Metodika a Technika Informaci* (Methodology and Technique of Information), No. 10. 1964, pp. 1-19. In Czech.

The author of this article worked for more than ten years in the Technical and Economic Information Center of Czechoslovakia and was in charge of processing information on metallic materials and their alloys. Aspects considered were properties, treatment, and applications of such materials. Experiences, gained by the author in the practical application of the UDC category 669 in classifying such materials, and suggestions for modifying the system are presented. (author)

10/65–226 **Statement of Types of Classification Available to New Academic Libraries, Report of the Classification Committee, RTSD Cataloging and Classification Section, May 15, 1964.** 9(1): pp. 104–110. Winter 1965. LRTS

A committee of the ALA Cataloging and Classification Section reported in May 1964 on the comparison of classification systems which might prove most adaptable and most durable over a long period of time. Does not consider the problems of changing from an existing classification. It proceeds to set forth the characteristics of such major classifications as Dewey, LC, Bliss, Rider, Reader-Interest, Universal Decimal, Colon, and Faceted Classifications. Pays attention to the Dewey and Library of Congress on such points as comprehensiveness, flexibility, complexity, indexing, browsability, notation, number of classes, interpretation of use, systems of subdivisions, revisions, and particular attention to the effect of the size of collections, the question of centralized cataloging, and relative costs, with a tentative comment on ease of application to a mechanized system. (D.C.W.)

Subject and Linguistic Analysis – General

10/65–227 **On Rules of Russian Conjugation: Preliminary Information** (O Pravilakh Russkogo Spryazheniya: Predvaritel'noye Soobshchensive), Morris Khalle. 1963, 20p., AD-436 075, Div. 32, Massachusetts Inst. of Tech., Cambridge. NSF Grants G16526, NIH MH 0473703, Unclassified report. Text in Russian. Reprint from American Contributions to the Fifth International Congress of Slavists, Sofia, pp. 113-132, 1963, Moulton and Co., The Hague. (Copies not supplied by DDC.)

The paper consists of a detailed examination of the Russian conjugation. It is shown that considerable simplifications can be achieved if the rules which describe the different forms are made to apply in a specific order. Further economies can be realized if the Immediate Constituent (IC) structure of the forms is taken into consideration, so that a part of the ordered rules is applied in cyclical fashion to successively larger and larger constituents beginning with smallest IC and ending with the IC that includes the entire form. It is shown in passing that the proposed rules account in a thoroughly natural fashion for certain other morphophonemic processes of the language such as the derivation of the adjectival forms of comparison. The appendix gives in outline form a proposed treatment of the derivation of imperfective verbs from prefixed perfective verbs. (author)

10/65–228 **Natural Language in Computer Form,** Martin Kay and Theodore Ziehe. AD-456 948, Div. 30, Rand Corp., Santa Monica, Calif., Memo. no. RM 4390 PR, Feb. 65, 81p. Contract AF49 638 700, Unclassified report.

This memorandum describes a scheme for recording text in computer-usable form in such a way that all meaningful typographical distinctions are represented in a standard way. Provision is made for texts in different languages and different alphabets and for subsidiary material such as parallel translations and comments of interest to users and librarians. The basic set of encoding conventions is indefinitely extensible to accommodate new kinds of material. Very large bodies of data require special facilities, and these have been provided by embedding the text encoding scheme in a general file maintenance system. Computer programs are described which simplify conversion of text from these various sources into the standard format. The final section discusses the problem of printing text which has been recorded in the standard format and describes a flexible program for doing this. (author)

10/65–229 **Quarterly Progress Report No. 23, Prepared for National Science Foundation, 1 November 1964, 31 January 1965,** W. P. Lehmann, Director, E. D. Pendergraft, Asociate Director, Grant NSF GN-308, Linguistic Research Center, The University of Texas, Box 7247, University Station, Austin, Texas 78712, LRC 65 NSF-23, January 1965.

Progress is reported in the development of two generalized computer systems: the first designed to implement automatic translation of languages, the second to support basic research in linguistics. The systems complement each other in that basic programs prepared for each are applicable to, and needed in, the other. The common system design contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first two sections are now operational. In the third, the syntactic translation algorithm was completed during the quarter. Its operational status was demonstrated by the production of a German-English translation of six paragraphs. Research in automatic classification, information retrieval and techniques of data acquistion are also reported.

10/65–230 **Results of Classifying Documents with Multiple Discriminant Functions,** J. H. Williams 15 Mar 65, 31p. Research progress rept., AD-612 272, Div. 32, CFSTI Prices: HC$2.00, MF$0.50. Contract Nonr445600, Unclassified report. Delivered at the Symposium on Statistical Association Methods for Mechanized Documentation, sponsored by the National Bureau of Standards.

An important, but frequently underemphasized step in the classification process is the selection of attributes. In classification problems of mutually exclusive assignment, a set of attributes is selected to represent the category. For information retrieval applications the assumption of mutually exclusive categories may not hold. Therefore, the problem of the selection of measurable attributes to represent the categories becomes more acute. Discriminant analysis appears to offer a solution not only to the selection of attributes problem, but also to the document relevance problem. In the selection phase it provides a method of selecting a set of attributes whose ratio of among-category variance to within-category variance is largest. In the actual classification process a distance measure can then be employed to determine the degree of relevance of a given document with respect to each category. Classification experiments have been conducted on 794 Solid State abstracts. Classification accuracies up to 90 percent were achieved using the discriminant procedures.

10/65–231 **Some Observations on the Performance of EJC Role Indicators in a Mechanized Retrieval System.** F. W. Lancaster. SPLB, *55*(10), pp. 696-701 (Dec. 1964). Problems are discussed which arise in analyzing information for retrieval from the

use of role indicators keying an item for example, as to whether it is active or acted upon. A study at Wright Patterson Air Force Base to determine relevance ratios of material recalled or missed, among other findings, showed that the addition of roles to links reduced recall by 5.2% while improving relevance by only 2% while more than doubling the time for programming and searching and the cost of input. It is suggested that the use of roles be dropped. (MCP)

10/65–232 **Computer Analysis of the Nuclear Test Ban Treaty.** R. A. Langevin and M. F. Owens. *SCIE 146*, pp. 1186-1189 (1964). Reports the results of experiments undertaken to determine the applicability of a computer for automatic syntactic analysis to the systematic discovery of ambiguities in textual material. *(K.G.)*

10/65–233 **A Matter of Syntax.** Ralph A. Lewin. *SCIE 147*, pp. 357-8, (1965). Gives an example of a particular syntactical dilemma and suggests various solutions. *(K.G.)*

10/65–234 **A Note on the Pseudo-Mathematics of Relevance.** Mortimer Taube. AMDO, vol. 16, no. 2, April, 1965. pp. 69-72.

Recently a number of articles, books, and reports dealing with information systems, i.e., document retrieval systems, have advanced the doctrine that such systems are to be evaluated in terms of the degree or percentage of relevancy they provide.

Although there seems to be little agreement on what relevance means, and some doubt that it is quantifiable, there is, nevertheless, a growing agreement that a fixed and formal relationship exists between the relevance and the recall performance of any system. Thus, we will find in the literature both a frankly subjective notion of relevance as reported by individual users, and equations, curves, and mathematical formulations which presumably provide numerical measures of the recall and relevance characteristics of information systems. This phenomenon of shifting back and forth from an admittedly subjective and non-mathematical term to equations in which the same term is given a mathematical value or a mathematical definition has its ancient paarallel in discussions of probability. One cannot, of course, legislate the meaning of a term. It all depends, as Alice pointed out, on "who is master," the user or the term. On the other hand, the use of a single term in the same document to cover two or more distinct meanings, especially when such a usage is designed to secure the acceptance of a doctrine by attributing to it mathematical validity which it does not have, represents a more serious situation than merely careless ambiguity.

Although there are a great many recent articles in the literature which deal with the notion of relevance, in the balance of this paper the demonstration of the illegal shift from subjective relevance as a reaction of a user to mathematical relevance as a property of systems will be restricted mainly to the work of the Cranfield Studies and the Arthur D. Little report, "Centralization and Documenation," since these two sources have given the widest currency to the pseudo-mathematics of relevance.

10/65–235 **Zur Terminologie des Starkegebietes.** 1. Mitteilung: Der Begriff der Starke. (Terminology of the Starch Field. 1st Communication: The Concept "Starch"). M. Ulmann. *Starke 16* (2), pp. 54-62 (1964). The word "starch" means different things to the botanist, the chemist, and the industrialist or merchant. The confusion is confounded by the near-synonymous use of "starch" and "starch grain" in the literature. A comprehensive team approach to arrive at standard definitions in the field of starch and starch derivatives is urged. *(C.L.B.)*

Application of Computers to Classification, Indexing, and Text Processing

10/65–236 **Some Parameters for Computational Stylistics: Computer Aids to the Use of Traditional Categories in Stylistic Analysis,** Sally Yeates Sedelow, Walter A. Sedelow, Jr., and Terry Ruggles. 1964, 21p. AD-613 376, Div. 32-, 30. System Development Corp., Santa Monica, Calif. Contract Nonr442700. Unclassified report. Pub. in mono. IBM Literary Data Processing Conference, Yorktown, N. Y. 1964. Proceedings pp. 211-29. (Copies not available to DDC or Clearinghouse customers).

"Computational Stylistics," the use of computers for a rigorous description and analysis of pattern attributes of text, is result of a desire to give contemporary technology focused upon language processing the benefit of analytical methods devised and used by literary critics and, at the same time, to give literary critics the benefit of tools provided by contemporary technology—i.e., the large digital computer and new programming procedures.

10/65–237 **Document Format Recognition.** Final report by Steven B. Gray. Jan 65, 471p. Rept. no. F-3060-1. Contract AF30 602 3273, Proj. 4599, Task 459902. RADC TDR-64-463. AD611 632, Div. 32, 30. CFSTI Prices: HC$4.75, MF$2.25. Sylvania Electric Products, Inc., Waltham, Mass. Unclassified report. Available copy will not permit fully legible reproduction. Reproduction will be made if requested by users of DDC. Copy is not available for public sale.

This study is primarily concerned with methods for analyzing the format of pages from technical journals, and means for automatically processing the textual and graphic material on these pages for input to computer which is to perform textual data processing functions, such as automatic language translation, automatic abstracting, automatic indexing, etc. This analysis and processing includes text-graphic separation, location of graphics, and textual analysis and recognition. The overall process is considered to be a Format Recognition and Analysis Program operating on a computer-controlled character recognition device. This study has resulted in general design techniques for Format Recognition and Analysis Programs applicable to any document which occurs with text and graphics intermixed. Two such programs have been completed, tested, and demonstrated for two technical journals, one Soviet and one U.S., and a third program has been outlined and partly written for another Soviet journal. It has been found that almost any journal can be programmed without serious difficulty, but new journals require substantially different programs. (author)

10/65–238 **Computational Research in Arabic,** Arnold C. Satterthwait. 1963, 9p. Unclassified report. AD-435 708, Div. 32, 30. Research Lab. of Electronics, Mass. Inst. of Tech., Cambridge. Reprint from Mechanical Translation, 7:2, pp. 62-70, Aug. 63. (Copies not supplied by DDC)

10/65–239 **Preparations for Setting-up a Bibliographic Index by Computer,** Helbich, J., *Racionalizace Prace Strojove Zpracovani Informaci v Knihovnach a Utvarech Tei* (Rationalization of Work and Mechanization of Information Processing in Libraries and Information Centers), No. 2, 1963, pp. 105-110. In Czech.

A method of compiling permutated title indexes, proposed by Luhn, is discussed. The key words, taken from the complete title, do not appear as isolated words. They appear in context. For example, the words preceding and following each key word are so arranged that, in most cases, the context in which the key word is used is clear. The index is code-linked to the basic bibliography. An innovation, which extends the framework of the Luhn method, is the use of titles in their original language, at least in Russian, English, and Czech. The National Elliot 803 B Calculator is the processing equipment used. (DOC INC)

10/65–240 **Machine Recognition of Human Language. III. Cursive Script Recognition,** Nilo Lindgren. *IEEE Spectrum*, v. 2, no. 5, May 1965, pp. 104-116. Outlines several approaches to mechanize the recognition of natural handwriting. (BATR)

10/65–241 **Computer Recognition of Handwritten First Names,** Frank N. Marzocco. IREL, v. EC-14, no. 2, Apr. 1965, pp. 210-217. A learning program incorporating a version of stimulus-sampling theory was prepared for a digital computer. Handwritten signatures coded in a 20 by 48 grid served as inputs to study the effects of parameter changes in the program. (BATR)

10/65–242 **Quarterly Progress Rept, No. 3, 1 Nov. 63-31 Jan. 64.** W. P. Lehmann, Feb. 64, 38p. Rept. nos. 19, 64P19.

Contract DA36 039AMC02162E. Unclassified report. AD-435 103, Div. 32, 30. Linguistics Research Center, Univ. of Texas, Austin.

Progress is reported in the development of two generalized computer systems: the first designed to implement automatic translation of languages, the second to support basic research in linguistics. The systems complement each other in that programs prepared for each are applicable to, and needed in, the other. The common system design contains three sections: one for control, a second for language data processing, and a third for linguistic information processing. The first two sections are now operational. In the third, Monolingual Recognition programs for performing lexical and syntactic analysis and display have been made operational. The programs, which have been converted to operate on the IBM 7040 as well as the IBM 7090 computer, are being tested with English, German, Russian, and Chinese language data. (author)

10/65–243 **Mechanized Indexing Studies of MSD Toxicity. Part I and Part II.** John O'Connor. Jan. 64, 72p. Contracts Nonr418300 AF49 638 1300. AD-436 523, Div. 32, 30, 16. OTS price $7.60. Institute for Scientific Information, Philadelphia, Pa. AFOSR 64 0682. Unclassified report.

Mechanized indexing possibilities are being studied by searching for computer rules which will duplicate indexing done by subject specialists for a pharmaceutical retrieval system, Toxicity (including side effects) is the first term studies. A random hundred documents, containing thirteen toxicity papers, is the first sample. Computer-generated word frequency lists for each paper aid the work. The much discussed keyword approach is tried first. A thesaurus group of likely toxicity keywords is derived from the retrieval system's indexing guide, a medical dictionary, and the papers in the sample. Frequencies and document places (occurrence in titles, summaries, headings, etc.) of keywords, and differing weights for keywords, are used to try to reduce keyword overassigning of toxicity. Frequencies do not appear to help. Weighting helps some. Occurrence in summaries helps more. No keywords occur in titles. The Appendices (bound separately) include a bibliography of all sample papers containing keywords. Each reference is accompanied by a brief summary of the paper, and brief summaries of the passages containing keywords. Another Appendix gives for each paper its word length and the identity, document places, and frequencies of the keywords it contains. Another Appendix contains the full texts or excerpts of the five toxicity papers which appear to contain no keywords. (author)

10/65–244 **Mechanized Indexing Studies of MSD Toxicity, Part II.** John O'Connor. 1964, 71p. Contracts Nonr418300; AF49 638 1300. Unclassified report. AD-437 868, Div. 32, 16. OTS price $7.60. Institute for Scientific Information, Philadelphia, Pa. Unclassified report.

A general procedure is described for using the meaning of a paper as a guide in looking for mechanized indexing rules. The procedure involves two basic concepts: (1) A paper saying something about T (toxicity, or any other index term being studied); and (2) an expression in paper being evidence that the paper says something about T. Neither of these notions is completely well-defined, but they may be sufficiently intersubjective to help find useful mechanized indexing rules. The application of the procedure to the toxicity papers in the sample described in Part I leads to the thesaurus keywords which were obtained more intuitively in Part I. It also leads to nine additional keywords, which function almost as well for mechanized indexing purposes. The general procedure leads to many other expressions which are also toxicity clues, but which cannot be used directly for mechanized indexing, because they are unlikely to recur in other toxicity papers. Study of these expressions suggests their generalization to expression forms, containing variables. The possible values of the variables are defined for computer use by lists of substance-contact words and disorder words. (author)

10/65–245 **Automatic Indexing: a State-of-the-Art Report.** Stevens, Mary Elizabeth. U. S. Dept. of Commerce, National Bureau of Standards, Washington, D. C. Issued March 30, 1965. (NBS Monograph 91) 220 p. USGPO price $1.50.

"A state-of-the-art survey of automatic indexing systems and experiments has been conducted by the Research Information Center and Advisory Service on Information Processing, Information Technology Division, Institute for Applied Technology, National Bureau of Standards. Consideration is first given to indexes compiled by or with the aid of machines, including citation indexes. Automatic derivative indexing is exemplified by keyword-in-context (KWIC) and other word-in-context techniques. Advantages, disadvantages, and possibilities for modification and improvement are discussed. Experiments in automatic assignment indexing are summarized. Related research efforts in such areas as automatic classification and categorization, computer use of thesauri, statistical association techniques, and linguistic data processing are described. A major question is that of evaluation, particularly in view of evidence of human inter-indexer inconsistency. It is concluded that indexes based on words extracted from text are practical for many purposes today, and that automatic assignment indexing and classification experiments show promise for future progress." (author)

10/65–246 **Automatic Classification and Indexing, for Libraries?** Donald V. Black. Winter 1965. LRTS 9 (1): pp. 35–52.

Reviews the major literature since Luhn's 1957 paper on mechanized encoding and searching of literary information. Comments on the practicality of automatic classification and indexing. Concludes with summary of his proposal for a system based on input of book title, table of contents, index if any printed in the book, introductory paragraphs that describe content, and possibly a short indicative abstract. (DCW)

10/65–247 **The World's Fair Machine Translator.** Dan M. Bowers and Miles B. Fish. CMPD, v. 4, Apr. 1965, p. 16 + 10 pp. This description of the Russian-to-English machine translation system includes the construction of the system, its dictionary organization, and its translating production operating mode. (BATR)

10/65–248 **Some Clumping Experiments for Associative Document Retrieval.** A. G. Dale and N. Dale. AMDO, v. 16, no. 1, pp. 5–9, 1965.

This paper describes experiments with a statistical technique that can be used to compute word associations useful for document retrieval purposes in a collection where documents are described by index words occurring in the text or in abstracts of the documents.

Although the experiments have been conducted with a small collection of documents, they have suggested methods that will permit the association technique to be applied in collections of an interesting size. This is important, since a number of alternative methods that have been proposed for computing word associations appear to present intractable computation problems that will deter large scale application.

The experiments described in this paper indicate the potential utility of clumping techniques as a basis for large-scale associative document retrieval. This conclusion is based on several considerations arising from the work described, and from work currently in progress:

1. Clumping associations appear to provide a valid basis for retrieval.

2. The efficiency of retrieval, particularly as measured by the total number of documents retrieved in response to a request, will not be degraded in large collections, as is the case with coordinate systems.

3. Clumping techniques can be applied to compute associations in very large sets of data. We are currently experimenting with computing techniques capable of rapid identification of clumps in data sets of the order of 1,500 index words, and expect to have programs capable of handling much larger sets (of the order of 10,000 index words) later this year.

10/65–249 **Evaluation of Machine Translations by Reading Comprehension Tests and Subjective Judgments.** Sheila M. Pfafflin, MTRN v 8, no. 8, Feb. 1965, pp. 2–8. Discusses the results of an experiment designed to test the quality of translations, in which human subjects were presented with IBM-produced machine translations of several passages taken from the Russian electrical engineering journal *Elektrosviaz*, and with human trans-

lations of some other passages taken from *Tele-communications*, the English translation of *Elektrosviaz*. Although the human translations generally gave better results than the machine translations, the differences were frequently not significant. (BATR)

10/65–250 **Document Association and Classification Based on L-Languages.** Jerry Sanders, JACM, v. 12, no. 2, Apr. 1965, pp. 249–253. Measures of document connectedness and association are defined based on a formal language structure. The properties and implications of these definitions are investigated and examples are given. The implications of the theory for automated document retrieval are discussed in conjunction with various extensions of the formal theory. (BATR)

10/65–251 **Sentence-For-Sentence Translation: An Example.** Arnold C. Satterthwait, MTRN, v. 8, no. 2, Feb. 1965, pp. 14–38. A computer program for the mechanical translation into English of an infinite subset of the set of all Arabic sentences was written and tested. Presents a generalized technique for thorough syntactic parsing of sentences by the immediate constituent method, a generalized structural transfer routine, and a consideration of the elements which must be included in a statement of structural equivalence with examples drawn from such a statement and the accompanying bilingual dictionary. (BATR)

10/65–252 **Machine Recognition of Hand Printing.** R. J. Spinrad. IFCN, v. 8, no. 2, Apr. 1965, pp. 124–142. The problem of pattern recognition as applied to line drawings was studied, and a generalized method developed which is based on a hierarchical sequence of classification stages. Recognition proceeds in two steps: the individual pen strokes are abstracted into straight line segments; then the normalized pattern of lines is identified with one member of the possible pattern set. The identification of hand printed letters was correct for 93.5% of the unknown samples. The operation of the entire system was simulated on a general purpose digital computer. (BATR)

10/65–253 **Improvements in a Permuted Title Index.** Hugh E. Voress. AMDO, vol. 16, no. 2, April, 1965, pp. 97–100

Improvements in a computer-prepared, permuted-title index are described. The improvements include adding words to present titles or synthesizing new titles when insufficient keywords are provided in the title, positioning singular and plural words and word variants so that one follows the other, identifying and adding cross-references where needed, and exceptional efforts in removing words that have no value for information retrieval. The index was prepared for a radiobiology bibliography.

Design, Testing, and Evaluation of Information Systems

10/65–254 **An Investigation into the Application of Data Processing to Library Filing Rules.** Don S. Culbertson, Louis A. Schultheiss, Arlon Sieve, Donald Boone, December 5, 1962, PB-164441, $.60. A joint endeavor by the University of Illinois Library, Chicago Undergraduate Division, Navy Pier, Chicago, Illinois, and the Burroughs Corporation.

This report is the result of a study to determine what problems would be involved in converting the present Card Catalog of a library to a book catalog which would be arranged by and printed on Data Processing Equipment. Questions pertaining to the initial conversion of present files, the implementation of present (A.L.A.) filing rules to a data processing system, the complexity of the programs required, and the type of data processing system most applicable have been answered.

The objectives set down at the beginning of the University Library Information Systems Project of which this is a continuing part are:

1. To automate, through an analysis of the flow of an item through the cataloging department, the present clerical operations of the library cataloger.
2. To determine, and incorporate in the design of the input media, the types of information contained on a catalog card so that subsequent retrieval of this information would be possible.
3. To automate the present filing operation according to specified filing rules.
4. To retain, as output from a data processing system, the full documentation presently found on the catalog card.
5. To produce, as output from a data processing system, printed lists which would replace the present card catalogs.

10/65–255 **Mathematical Theories of Relevance with Respect to the Problems of Indexing. Report No. 2: An Algorithm for Document Characterization.** Donald J. Hillman, March 1965. Center for the Information Sciences. Lehigh University, Bethlehem, Pa. NSF grant GN–177. 56 pp.

A distinction is made between document retrieval systems and "fact retrieval" systems, and it is stipulated that for the former the index terms should be the names of the topics dealt with by the documents in the system collection. Such index terms are called "document characteristics."

A document is then regarded as a complex assertion, and the problem of discovering its characteristics is defined to be that of isolating the referring expressions in the components of the complex assertion.

This is first done for simple English sentences by associating each of them with one of two canonical schemata. The referring expressions for each canonical schema are then identified.

It is next shown that the type of reference discernable in simple sentences is preserved when such sentences are transformationally combined to produce complex sentences. It is then shown that the referential function of a complex sentence may be investigated in terms of the referring expressions of its simple components.

Two methods of sentence reduction are examined for this purpose, *viz.*, the derivation of micro-sentences and a kernelization program. Kernels are inefficient for document characterization purposes. Hence, an algorithm is constructed which operates on kernels to form certain micro-sentences called "assertive components." This algorithm together with a method for weighting the referring expressions of assertive components provide the means for assigning characteristics to any given document. The characteristics accurately denote the topics about which assertions have been made in the document, and the weighting of the characteristics supplies a means for assessing how much of the document's content is taken up with a discussion of those topics.

10/65–256 **Review of a Report on the Aslib-Cranfield Test of the Index of Metallurgical Literature of Western Reserve University.** Alan M. Rees. AD–435 142, Div. 32, 17. Center for Documentation and Communication Research, Western Reserve Univ., Cleveland, Ohio. 16 Oct. 63, 32p. Grant NSF G10338. Unclassified report.

The application of the Aslib-Cranfield testing technique to the Western Reserve University Index of Metallurgical Literature is reviewed from the WRU viewpoint in the light of the experimental results derived from more than one hundred searches. The final report is analyzed in some detail with reference to relevance and recall code thesaurus, role indicators and punctuation levels. It is concluded that the search results were greatly influenced by the lack of recourse to the questioner during question analysis. (author)

10/65–257 **The Role of Graphic Display of Concept Relationships in Indexing and Retrieval Vocabularies including a Thesaurus of Documentation Terms.** L. Rolling. EUR 2291. Brussels, March 1965—29 pages—7 figures.
European Atomic Energy Community—EURATOM, Directorate "Dissemination of Information," Center for Information and Documentation—CID. Paper presented at the FID/CR International Study Conference on Classification Research—Elsinor (Denmark)—September 14-18, 1964.

Emphasizing the importance of graphic techniques for information exchange in general, and for documentation in particular, the author describes the preparation of terminology charts for a keyword thesaurus of documentation terms, and the use of arrowgraphs in the Euratom Thesaurus of nuclear energy.

The usefulness of this graphic display of concept relationships for thesaurus updating, indexing and retrieval is established.

Future developments, particularly in the fields of multilingual documentation and automatic documentation, are assessed.

10/65–258 **A Programming System for Coordinate Retrieval and File Maintenance.** A. T. ten Broeke. Prepared for National Science Foundation Grant NSF GN–308, Linguistics Research Center, The University of Texas. LRC 65 TS–3, April 1965.

A general information retrieval programming package written in Fortran IV is described. It provides a capability of retrieving information on the basis of any logical sequential set or sets of descriptors meeting the specifications and constraints of the program. The package also includes file creation and maintenance routines.

10/65–259 **Inefficiency of the Use of Boolean Functions for Information Retrieval Systems,** J. Verhoeff, W. Goffman, and Jack Belzer. 1960, 3p. AD–612 853, Div. 32, 15, 30. Contract AF49 638 357, AFOSR 1775. Unclassified report.

Pub. in Communications of the ACM (U. S.) v. 4, no. 12, pp. 557–9, Dec. 61. (Copies not available to DDC or Clearinghouse customers.)

It is shown why, in general, boolean functions are not applicable in information retrieval systems.

10/65–260 **Techniques for Thesaurus Organization and Evaluation.** C. T. Abraham. *Parameters of Information Science.* pp. 485–497. (See 4/65–1.)

For a vocabulary list with synonymy and hierarchy as the two kinds of pairwise relationships defined between terms, techniques have been found which will: (a) group together terms which are mutually synonymous, (b) provide consistent hierarchies between terms when they exist and (c) detect inconsistencies in the specification of pairwise hierarchical relationships. In addition, methods of comparing two thesauri for consistency using overlapping vocabulary lists have been developed.

10/65–261 **Graphic Data Processing,** Irving Abzug, DTMN 11, No. 1, pp. 35-37 (January, 1965).

The Graphic Data Processing System allows the user to communicate with a computer in almost the same way he talks to his colleagues. This computer-controlled system is intended to help engineers, designers, and businessmen work directly with graphics—charts, sketches, and drawings. The major elements in the Graphic Data Processing configuration include an IBM System/360 processor, visual display consoles, a 35-mm film recorder, and a 35-mm film scanner.

10/65–262 **Demonstrating Remote Retrieval by Computer at Library/USA.** Joseph Becker. *ALBL*, vol. 58, no. 9, October 1964, pp. 822-4.

The retrieval of information at Library/USA at the New York World's Fair using a Univac 490 Real-Time computer is described. The computer consists of a console, a control unit, a central processor a 1004 card processor and punch, a communications subsystem, and a random-access Fastrand mass-storage magnetic drum. Each of the six reference desks has a Uniset device from inputting inquiries directly to the computer and three output High-Speed printers are within reach of the reference librarian. A seven-hundred-word print-out can be provided for any request in less than four seconds. The 490 is able to communicate within a Univac 1004 anywhere in the world where data transmission lines are available. The advantages of the Real-Time computer over conventional computers are pointed out. The time between successive operations in the Real-Time computer is very much smaller so the processing of information is speeded up greatly. Storing the data on a magnetic drum which revolves continuously means that information may be located directly; the drum is referred to as a random-access memory. To find the same information on a magnetic tape takes much longer because the tape must be examined serially. (JDOC)

10/65–263 **Optimising Information Retrieval.** H. Bouman, REDO, vol. 32, no. 2, mai 1965, pp. 46–53.

Described are methods for optimal storage and retrieval of information using information carriers with limited capacity, in particular punched cards. One method, termed "scrambling," consists of grouping large numbers of randomly selected classes under a small number of notations. The second proposed method, termed "dechaotizing," involves the storage of the information carriers in a fixed order derived from the code pattern. (REDO)

10/65–264 **Some Prerequisites to Co-operative Cataloging.** Ritvars Bregzis. *CRLI*, vol. 25, no. 6, November 1964, pp. 497-500.

Two basic alternative co-operative cataloguing schemes for non-American material have been suggested: centralized cataloguing in which the library first acquiring a given item forwards it to a proposed centre; dispersed and co-ordinated cataloguing in which the proposed centre functions as a clearing house receiving and distributing requests for cataloguing which is to be performed by the library first acquiring the given items.

The second suggestion requires a smaller capital investment because the larger research libraries would do most of the cataloguing, but it has the disadvantages of giving these libraries extra work and of producing non-standard entries.

Centralized catalogues seems to be more promising. However, lending the copies for cataloguing would produce problems. As many research libraries will have automated acquisitions control, the records system must be compatible with mechanized records. Also the centre has to design rules for entry classification, and subject headings which are acceptable to the participating libraries.

There is a growing number of libraries requiring machine-usable catalogue records. Considerations of bibliographic service and economy require these to be based on a logical and uniform system of entry, systematization of subject matter, and systematic terminology control. Existing rules of entry, most classification schemes and subject heading systems incorporate logical and structural deficiencies that have been found by experience to prohibit effective automation of bibliographical control. The classification scheme required would not be linearly heirarchical; it would function as a heterogeneous logical complex of various aspects. Such a scheme would in addition convert the conventional system of subject headings into a conceptual system of terminology which could be used in conjunction with the concept classification system and serve as a terminological index to that system. This would permit direct conversion from terminology to classification or vice versa. A standard bibliographic format that leads itself to direct recording and reading by data processing equipment is required for any serious effort to establish co-operative processing or bibliographic information transmission on an automated basis. (JDOC)

10/65–265 **Redirection of Research into Associative Retrieval.** Edward C. Bryant. *Parameters of Information Science.* pp. 503–505. (See 4/65–1.)

The March 1964 symposium on statistical association sponsored by NBS and ADI indicated a stalemate in development of usable techniques. It is suggested that research needs be redirected toward á study of mathematical structures and the development of models and procedures which are based upon well defined objectives. It is suggested that large scale computer projects, in the absence of structural analysis, be viewed with skepticism.

10/65–266 **NASA Search System Analysis Sheet.** W. T. Brandhorst, P. F. Eckert. AMDO, vol. 16, no. 2, April 1965, pp. 124-126.

Since May 1962, the Scientific and Technical Information Facility operated for NASA by Documentation Incorporated has carried out, edited, and delivered to selected NASA contractors nearly 700 literature searches. These searches are not necssarily restricted to the collections of the Facility, which now number well in excess of 100,000 documents. The open literature materials announced by *International Aerospace Abstracts* have been included on the Facility's tapes since January 1963. A machine search by the Facility, therefore, can be said at the very least to provide complete coverage of items in both *Scientific and Technical Aerospace Reports (STAR)* and *International Aerospace Abstracts (IAA)*, the two principal abstract journals in the aerospace field.

A NASA Search System Analysis Sheet has been designed to gather statistics which would otherwise be widely dispersed in files. It is serving its purpose at the Facility and the general consensus is that an examination of these forms provides a real "feel" for how the machine searching activity is going as a whole, and how individual analysts are progressing in particular. The form is not only passed between Departments but between analyst

e method of protection against humidity is the strict ce of a temperature of 16–18 degrees Centigrade around and repeated ventilation (3–5 times daily) of all rooms 0–15 minutes. The destructive work of microorganisms and insects could be controlled by frequent cleaning of floors and desks. (DOC INC)

10/65–285 **Searching Titles by Man, Machine, and Chance.** Vaun A. Newill and William Goffman. *Parameters of Information Science.* pp. 421–423. (See 4/65–1.)

An experiment was conducted under the assumption that there is no difference in effectiveness when searching titles by man, machine or chance.

A file of 210 titles of articles in a subarea of the field of diabetes was searched by (a) an expert in the area, (b) a computer, and (c) the Monte Carlo method.

Questions to which answers were known *a priori* represented active research in the field. The various outputs were evaluated in accordance with methods developed in the Comparative Systems laboratory, Center for Documentation and Communication Research, of Western Reserve University.

10/65–286 **14983 The Accuracy of Federal Academic Library Statistics,** Eli M. Oboler. *CRLI,* 25 (6) November 1964 pp. 494–496.

Examination of American academic library statistics for 1962–63 reveals many imperfections. The figures for "volumes" and "volumes added" give particular cause for concern. For example, one library, in 1961–62, had a stock of 50,000v. It reported 5,000v. added during 1962-63, and its total stock at the end of that year as 100,000v. Many libraries give figures which are obviously estimates, without indicating the fact. There is lack of agreement between libraries as to whether microtexts and government publications should be included. There is a need for clarification of what is wanted and for a better understanding by respondents of what is asked for. The US Office of Information should publish only those statistics that make sense. (LSCA)

10/65–287 **A Brief Survey of Topological Representations.** Ascher Opler. *Parameters of Information Science.* pp. 499–502. (See 4/65–1.)

Many different techniques have been developed for representing a linear graph (network of nodes and branches) so that it will be useful for computer manipulation. Methods of representation include storage of diagrams directly, use of tables, lists and matrices. Among useful manipulations are determination of partial or complete isomorphism, transformation between representations, input and output conversion, generation of new graphs, etc.

10/65–288 **Document Compaction by Variable Length Encoding.** C. V. Ramamoorthy. *Parameters of Information Science.* pp. 507–513. (See 4/65–1.)

Discusses the use of machine-encoded abstracts to facilitate the storage and retrieval of documents. This encoding, by reducing the memory access time, minimizes the storage space required on magnetic tapes, drums, and disks and generally facilitates a faster search and retrieval of documents.

10/65–289 **The Principle of Minimum Control.** J. Albert Sanford and H. Edmund Stiles. *Parameters of Information Science.* pp. 425–426. (See 4/65–1.)

This paper is a statement of a basic principle which is frequently overlooked by designers of information retrieval systems. The major premise set forth, holds that in documentation that indexing scheme is best which exerts minimum control commensurate with real retrieval needs.

10/65–290 **The Comparative Systems Laboratory at Western Reserve University: Progress Report on Operational Procedures.** Tefko Saracevic. *Parameters of Information Science.* pp. 427–431. (See 4/65–1.)

In the first phase of the Comparative Systems Laboraroy (CSL), comparative experiments are being conducted to assess the influence of origin of inputs (title, abstract, full text), indexing languages, coding and format of output on the effectiveness of information retrieval systems. An attempt is made to hold constant such variables as acquisition, organization of the files, question analysis and search strategies.

The models tested, measures developed and proposed studies are briefly discussed. The actual operational procedures applied in the CSL are outlined in considerable detail.

10/65–291 **Answering English Questions by Computer: A Survey.** R. F. Simmons. *CACM,* v. 8, Jan. 1965, pp. 53-70. Fifteen experimental English language question-answering systems which are programmed and operating are described and reviewed. The systems range from a conversation machine to programs which make sentences about pictures and systems which translate from English into logical calculi. Systems are classified as list-structural data-based, graphic data-based, text-based and inferential. Principles and methods of operations are detailed and discussed. Concludes that the data-base question-answerer has passed from initial research into the early developmental phase. (BATR)

10/65–292 **An analysis of Factors Causing Irrelevant Answers to Machine Literature Searches and Proposed Solutions to the Problem.** Barbara B. Shaffer. *Parameters of Information Science.* pp. 433–436. (See 4/65–1.)

This paper presents an analysis of factors causing the disparity between computer responses to machine literature searches and answers actually sent to subscribers. Disparity sources are classified broadly as those inherent in the system, and improper use of the system on input and output levels.

Solutions are proposed for reducing irrelevant responses and eliminating manual evaluation.

10/65–293 **Information Networking.** Norbert Stahl. MEEN, v. 86, Dec. 1964, pp. 34-37. Network principles can provide a new way to analyze, simplify, and implement the flow of information, both technical and administrative. (BATR)

10/65–294 **Matching of Question and Answer Terminology in an Education Research File.** Jean Tague. AMDO, v. 16, pp. 26–32, 1965.

The hypothesis that relevant answers in a computer output may be distinguished from peripherally relevant and nonrelevant answers by the relative frequency with which question words and other related terms match words in the title, abstract, and machine index to these answers is investigated, using a sample of 14 questions searched against a pilot information file for education research. The frequencies of (1) significant words of the question, (2) words related thesaurally to the question words by the semantic code, and (3) words added to the original statement by the question analyst as answer index terms are used to determine empirically a function indicating the indexing and evaluating effectiveness of each. Similarly, the frequencies of question words in titles, conventional abstracts, and machine indexes are used to compare the indexing and evaluating effectiveness of these three forms of index.

10/65–295 **Inventorying and Updating the Holdings in Technical Libraries, J. Trautman,** *Methodicke Letaky* (Methodological Pamphlet), No. 45, 1964, pp. 1–15. In Czech.

The principles of inventorying the holdings of technical libraries and the policy of new acquisitions are discussed. Different approaches are postulated with regard to books, periodicals, and special technical literature such as reports, patents, and company literature. Various criteria for evaluating the obsoleteness of material in libraries are analyzed. Methods and procedures for inventorying library holdings are described. (DOC INC)

10/65–296 **Effect of Verbal Context on Latency of Word Selection.** Anne M. Treisman. *NATU,* v. 206, Apr. 10, 1965, pp. 218–219. Discusses two classes of models for word retrieval systems: noninformational models and information-processing models. (BATR)

10/65–297 **Two Characteristics of Circulation and Their Effect on the Implementation of Mechanized Circulation Control Systems.** *College & Research Libraries,* vol. 25, no. 4, July 1964, pp. 285–91.

An attempt is made to evaluate the need to prepare punched book cards for all volumes in the library in order to have a mechanized loans record system. The alternative is to prepare a punched book card for each book that goes out on loan provided it does not already have a punched book card. The data used in this research is taken from an analysis of the book loans record at the Deering Library of Northwestern University. The factors considered are the length of time the book has been in the library and the last previous loans date. It is shown from the data that it is considerably cheaper to prepare cards for books as they go out on loan than for the whole collection. Data connected with the previous date of loan gives a better indication of the time and work required to implement a punched book card loans system than does the data connected with the length of time a book has been in the library. (JDOC)

10/65–298 **Morphological Analysis by the Method of Successive Limitations,** D. Varga, *Orszagos Muszaki Konyvtar es Dokumentacios Kozpont. Idoszeru Muszaki Dokumentacios Kerdesek* (Hungarian Central Technical Library and Documentation Center. Current Problems of Technical Documentation), No. 3/4, 1963, pp. 244–266. In Hungarian.

The paper describes a new computer method which enables simple and quick determination of the stem of declined or conjugated words and the grammatical interpretation of the suffix. By this method, certain grammatical forms are made to correspond with particular bits of the machine-word (e.g., case inflections of nouns which are formed by suffixes, or certain forms of a verb). Boolean vectors are assigned to the examined words. The vectors indicate the category to which the word-form belongs. The simultaneous occurrence of different properties furnishes a basis to narrow down the set of possible categories. This can be done quickly on the machine by logic multiplication. By successive application of this technique, the narrowest partial set to which the word belongs is determined. For the morphological analysis of a word, the Boolean vectors that are required may be defined by the (a) type of letters appearing in the suffix, (b) number of letters in the suffix, (c) information taken from a dictionary. (C.B.L.)

10/65–299 **Dissemination of Information.** I. A. Warheit. *LRTS*, 9 (1): pp. 73–80. Winter 1965.

Briefly describes the use of and techniques for creating common computer-prepared dissemination tools.

10/65–300 **Design of an Experimental Multiple Instantaneous Response File.** E. L. Younker, C. H. Heckler, Jr., D. P. Masher, and J. M. Yarborough, Stanford Research Institute, Menlo Park, California. Proc. Spring Joint Computer Conf. 1964. pp. 515–28.

An experimental model of an electronic reference retrieval file in which all file entries are interrogated simultaneously has been designed and constructed. The experimental model is designed to store and search on a file of indexes to 5,000 documents. A document index consists of a decimal accession number and up to eight English word descriptors that are closely related to the contents of the document. The vocabulary required to describe the documents is held in a machine dictionary that has a design capacity of 3,000 words. In the model delivered to the sponsor, Rome Air Development Center, the storage capacity is only partially used. The specification for the delivered model calls for the storage of approximately 1,100 documents that were selected from the ASTIA (now DDC) Technical Abstract Bulletin and of the vocabulary needed to describe them (about 1,000 words). The document indexes and the dictionary words are stored in wiring patterns associated with arrays of linear ferrite magnetic cores.

A search question, consisting of one to eight descriptors in their natural English form, is entered by means of an electric typewriter. During entry of the search question, the dictionary magnetic store is interrogated by the alphabetic code of each search word. If a word is not contained in the dictionary, it is automatically rejected. After all words of the search question have been entered, the document magnetic store is interrogated by the search question in superimposed code form. The comparison between the search word and the document indexes is made for all documents simultaneously and the machine instantaneously determines if any documents in the file include the search question. If there are none, the machine indicates visuall[illegible] is no response. If there is at least one, the machine c[illegible] number of responding documents and displays this number, [illegible] it types out the indexes of all responding documents on the s[illegible] typewriter that was used to ask the question.

10/65–301 **Further Implications of the Distribution of Index Term Usage.** Eugene Wall. *Parameters of Information Science.* pp. 457–466. (See 4/65–1.)

Equations for approximately predicting vocabulary size and typical index (and search) term usage as functions of index size are presented. Computer methods for developing more exact equations are described. Improved index evaluation techniques may result from these concepts.

10/65–302 **An Application of a Type of Matrix to Analyze Citations of Scientific Papers.** Eri Yagi. AMDO, v. 16, No. 1, pp. 10-19, 1965.

A scientific paper has a series of citations belonging to the previous literature. In order to determine the linkage pattern of scientific papers through their citations, an analysis of a group, whose papers are frequently cited by themselves, is useful. It is possible to investigate the internal connections among papers in such a group by examining their citations. The use of a type of matrix for the analysis was suggested by Derek Price while the techniques of the matrix analysis had been developed for the study of group structure in the fields of sociology and psychology.

The main purpose of this study is to investigate the measurements by which the internal connections of a group of scientific papers are measured.

10/65–303 **Comments on the 1964 ADI Meeting.** Carlos A. Cuadra. AMDO, vol. 16, no. 2, April 1965, pp. 105–112.

A group of attendees and program participants was surveyed after the 1964 meeting of the American Documentation Institute to provide ideas and guidance for future ADI meetings. The responses indicated that attendees view the quality of the technical program as the key to a successful meeting; that a four-day meeting length meets with wide approval; that, if carefully planned, concurrent plenary sessions might be acceptable; that improvements in scheduling are required to permit greater participation in Author Forums; and that conference proceedings greatly enhance the value of the meeting and should be distributed well before the meeting. The many suggestions and criticisms that were offered are being studied in relation to the next regular ADI meeting in 1966.

10/65–304 **Mechanised Information Retrieval in the United Kingdom.** Harry East. *TVF. Teknish-Vetenskaplig Forskning,* v. 36, no. 2, 1965, pp. 64–66. (BATR)

The use of tape typewriters in libraries in preparing research and development abstracts, catalogue cards, and accession lists is discussed. Examples are given of keyword indexing systems and coordinate indexing, as well as investigations made in the Aslib Research Department on computerized systems.

10/65–305 **Indexing, Automation and the User's Requirements.** Robert S. Hooper. *Parameters of Information Science.* pp. 471–475. (See 4/65–1.)

This paper discusses a vocabulary of descriptors which have been used for indexing literature related to documentation and information science. The descriptors, and their categories were derived from a study of requests for information from this literature. Efforts to automate this indexing operation are examined and preliminary results are noted. This paper emphasizes that the starting point for a modern indexing and retrieval system must remain with the traditional study of the user's requirements.

10/65–306 **A Konyvtari Munka Osszehasonlito Merese Standardizalas Utjan** (A comparative measurement of library work with standardization) Jozsef Kovacsics and Katalin Duxne Nagy. *Magyar Konyvszemle,* 80 (3) July–September 1964, pp. 225–238.

The authors criticize the traditional methods of library statistics which compare the libraries on the basis of absolute figures and percentages relating to libraries of different size. By

American Documentation

VOLUME 16

1965

Published by

THE AMERICAN DOCUMENTATION INSTITUTE

CONTENTS

No. 3, July 1965

No. 4, October 1965

LITERATURE NOTES—Please enter my subscription for one year commencing with the January 1965 issue. At $4.00 per year, payment is enclosed ☐

bill me. ☐

Name ____________ Title ____________

Address ☐ Business ☐ Home ____________

City ____________ State ____________ Zip ____________

Your Firm Name ____________

Literature Notes Separate

The Literature Notes section of American Documentation is now available as a separate publication.

Literature Notes will continue to be an important section of American Documentation. However, popular demand has persuaded ADI to offer this separate quarterly subscription as an offprint.

On a Calendar year basis, the subscription price of Literature Notes is $4.00 per year, commencing with the January 1965 issue.

Return the coupon above. Payment with the order is requested.